Intelligent and Affectively Aligned Evaluation of Online Health Information for Older Adults

Julie M. Robillard
National Core for Neuroethics
Djavad Mowafaghian Centre for Brain Health
Division of Neurology, Department of Medicine,
University of British Columbia, Vancouver, BC

Areej Alhothali Sunjay Varma Jesse Hoey
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario

Abstract

Online health resources aimed at older adults can have a significant impact on patient-physician relationships and on health outcomes. High quality online resources that are delivered in an ethical, emotionally aligned way can increase trust and reduce negative health outcomes such as anxiety. In contrast, low quality or misaligned resources can lead to harmful consequences such as inappropriate use of health care services and poor health decision-making. This paper investigates mechanisms for ensuring both quality and alignment of online health resources and interventions. First, the recently proposed QUEST evaluation instrument is examined. QUEST assesses the quality of online health information along six validated dimensions (authorship, attribution, conflict of interest, currency, complementarity, tone). A decision tree classifier is learned that is able to predict one criteria of the QUEST tool, complementarity, with an F1-score of 0.9 on a manually annotated dataset of 50 articles giving advice about Alzheimer disease. A social-psychological theory of affective (emotional) alignment is then presented, and demonstrated to gauge older adults emotional interpretations of eight examples of health recommendation systems related to Alzheimer disease (online memory tests). The paper concludes with a synthesizing view and a vision for the future of this important societal challenge.

Introduction

There are five million Americans currently living with Alzheimer disease (AD) or another dementia. The costs of dementia care are estimated at over $200 billion, and are expected to reach $1 trillion by 2050. Faced with this epidemic and fearing the devastating impact of dementia on autonomy and quality of life (Corner and Bond 2004), older adults and their families are increasingly seeking online resources about the prevention and treatment of AD. However, unregulated online resources may contain inaccurate information, and may present information in a misleading or emotionally mis-aligned manner. These denotative (cognitive) and connotative (emotional) elements combined may lead to important harms, such as weakened patient-physician relationships, increased burden on health care systems, increased susceptibility to fraud, and negative health decision-making. Limitations of online health resources have also been identified by users: they report difficulties in identifying and using appropriate information, and they may be harmed by over-consumption, for example by experiencing anxiety as a result of consulting online resources (Benigeri and Pluye 2003; Eysenbach et al. 2002; Libert 2015; Blank and Lutz 2016). Further confusion may arise when unregulated and potentially commercially-driven website content is represented as educational (Wolfe 2002; Robillard et al. 2015).

The emergence of interactive online self-assessments to screen for age-associated conditions such as AD and other dementias is introducing additional challenges to the online environment (Robillard et al. 2015). For example, the limited oversight of online tests may lead to inaccurate self-diagnosis and treatment with negative health consequences, and the adverse psychological impact of poor test results may be magnified by the absence of appropriate in-person support and counseling. These risks are particularly relevant when the consumers of the information are older adults with possible cognitive impairment. On the other hand, high-quality and emotionally aligned information help Internet and mobile users to feel empowered about their health, inform positive health decision-making, and promote positive experiences and interactions with health care professionals (Hartin et al. 2016; Norman et al. 2008).

As the popularity of online health information and self-assessments continues to rise (Giles and Newbold 2011; Lovett, Mackey, and Liang 2012; Ryan and Wilson 2008), it is imperative to develop broadly accessible and emotionally aligned tools to rapidly evaluate the quality of online resources and to deliver this information in a way that ensures the potential harms of these interventions are minimized.

Several instruments have been developed in response to the need for quality evaluation of online health resources. These instruments are typically comprised of several criteria for the evaluation of static online content and must be applied manually by the Internet user. Currently available evaluation instruments include generic assessments as well as tools targeted to a specific 1) health condition (Hsu and Bath 2008; Seidman, Steinwachs, and Rubin 2003); 2) aspect of a condition such as treatment (Charnock et al. 1999); or 3) audience (Moult, Franck, and Brady 2004; Jones 1999). Examples of these tools include the HON Code, the DISCERN and the LIDA instruments. For a review of existing evaluation tools, please see (Zhang, Sun, and Xie 2015). Limitations of existing tools include narrow scope of application (e.g., only for treatment information), length of application (e.g., 15 or more criteria), lack of quantitative results, making it difficult to compare different sources of information, and deficits in evaluations of the reliability and validity of the instruments themselves. To address these limitations, the QUEST instrument was recently developed and provides a reliable, quantitative and valid quality evaluation tool that can be rapidly applied to a broad range of health information (Robillard and Feng 2016). QUEST measures six aspects of the quality of online health information (authorship, attribution, conflict of interest, currency, complementarity, tone) and demonstrates high levels of inter-rater reliability as well as convergent validity with other established quality evaluation instruments. While QUEST is currently designed to be manually applied, it is ideally suited to automation.

Quality evaluation for interactive online resources such as self-assessments is more complex than for static online content as they involve more significant ethical and emotional dimensions. Self-assessments for AD provide diagnostic, prognostic and prescriptive elements, and hold the potential to directly intervene in a persons decision-making process, rather than simply inform. These direct and interactive interventions raise ethical and emotional issues such as consent (the permission to undergo a medical intervention in knowledge of the benefits and risks), privacy and confidentiality (the requirement to preserve the privacy of identifying personal and health information), framing of the results (the need to avoid features that can negatively impact mental health) and variable quality of customized advice delivery (the need to ensure accuracy and usefulness of the information for the end-user). As such, evaluation and feedback about interactive resources must take into consideration their affective and emotional impact (Robillard et al. 2015).

This paper presents automated methods for the analysis of both denotative and connotative aspects of two types of online health resources: 1) static websites containing articles about health; 2) interactive self-assessment tools. First, a method for classifying static web content for *complementarity*, or whether the page supports the patient-physician relationship, is demonstrated. The method uses a simple bag-of-words model and a decision tree classifier. In the second part, a social psychological theory of identity and emotional alignment is used to predict attitudes and behaviors in response to online self-assessments for dementia. Taken together, these lines of inquiry demonstrate how artificial intelligence can be applied to the urgent challenges created by the unregulated online health environment.

Methods

Automated quality analysis of online health information

QUEST instrument

The QUEST instrument is a seven-item tool that quantitatively measures six aspects of the quality of online health information using weighted criteria and yielding an overall quality score ranging from 0 to 28. Criteria assigned a weighting of 1, such as authorship, currency and complementarity, are generally present in high quality publications; however, their presence is not necessarily reflective of overall information quality. Criteria assigned a weighting of 3 (attribution, conflict of interest, tone), represent more reliable and consistent indicators of quality (Robillard and Feng 2016). Content validity was ensured by developing the six proxy measures based on past research in this field (Chumber, Huber, and Ghezzi 2015; Sandvik 1999; Silberg, Lundberg, and Musacchio 1997). The QUEST instrument was found to have excellent inter-rater reliability and convergent validity and has been applied to samples of online health information related to both treatment and prevention (Robillard and Feng 2016).

Automation of the instrument: Complementarity

This section describes the automation of the *complementarity*, C, criteria of the QUEST tool. This criteria distinguishes documents that show clear "Support of the patient-physician relationship" from those that do not, for example by including a statement such as "Talk to your doctor before you start any exercise program". The automation of this criteria is a first step towards the full automation of the QUEST tool.

A set of 50 webapges were selected at random from the full set of 308 that were sampled in (Robillard and Feng 2016), keeping an even split between pages displaying complementarity ($C = 1$) and non-complementarity ($C = 0$) as scored by (Robillard and Feng 2016). The webpages were manually checked, and broken links were replaced with another random choice with the same C label. The main content of each page was extracted manually, and this content is referred to as a "document" hereafter. Each document was then re-annotated for complementarity at the paragraph level. Paragraphs were marked with $C = 1$ if they contained a sentence that showed clear support of the patient-physician relationship. Each document was marked with $C = 1$ if *any* paragraph within it scored $C = 1$. The final set was comprised of 26 pages with $C = 1$ and 24 pages with $C = 0$.

Re-annotation was necessary for two reasons. First, because many pages had changed since the analysis of (Robillard and Feng 2016). This discrepancy in the data sets at two separate time points illustrates the dynamic nature of the online environment and further supports the need for automated tools that can provide a rapid assessment by the user at the moment of consultation. Second, the paragraph level features were found to be more effective for learning the classifier. This is due to an increase in the number of annotated data and to a greater specificity of the features in each paragraph.

The documents were then cleaned by removing stop words (e.g., "the", "and"), numbers, punctuation and by normalising to lower case. Each paragraph was then converted to a "bag-of-words": a feature vector giving the counts for each word used. A leave-one-out cross validation method was then applied as follows. For each document, a CART model (decision tree with best-GINI index splitting (Breiman et al. 1984)) was trained on the dataset formed by removing that document (on the remaining 49 documents). The decision tree learning method was used to seek the feature that splits the dataset into two parts such that the information content is maximized overall. The Gini impurity is used as the splitting criterion and splitting was stopped at a minimum impurity threshold of 0.1. The decision tree was then used to classify the left-out document as $C = 1$ or 0 by classifying each paragraph and taking the maximum (OR of all C values). This predicted value was compared to the document's annotated value. This process was repeated for every document, yielding 50 predictions.

Characterization of emotional alignment in online self-assessments

This section describes the characterization of affective and emotional alignment in the use of online self-assessments for AD using a social psychological theory called Affect Control Theory (ACT) (Heise 2007). First, the principles of ACT are described, followed by details of how this theory was applied to a sample of online self-assessments.

Affect Control Theory

Affect Control Theory (ACT) arises from work on the sociology of human interaction (Heise 2007). ACT proposes that social perceptions, behaviors, and emotions are guided by a psychological need to minimize the differences between culturally shared fundamental affective sentiments about social situations and the transient impressions resulting from the interactions between elements within those situations. This affect control principle has been shown to be a powerful predictor of human behaviour (MacKinnnon and Robinson 2014). While ACT is based in sociological evidence about human interactions with other humans, it is known that people also ascribe affective meaning to media (Reeves and Nass 1996), and to technological artifacts (Shank and Lulham 2016; Shank 2010). Thus, ACT predicts that when engaging with an online self-assessment for AD, a person will have an affective interpretation of the website/test and will behave in a way that is heavily influenced by this interpretation. For example, a test that appears to have been developed by a reputable physician may be responded to with trust and deference, whereas one that is presented as the product of a cultish guru may be met with ridicule and repudiation. These interpretations, and their subsequent associated behaviors depend heavily on the identity or personality of the individual taking the self-assessment. An anxious and meek person may be more susceptible to suggestions from a powerful looking doctor than a relaxed and confident person.

Both the fundamental affective sentiments and transients impressions can be represented as vectors in a three-dimensional affective space, the basis vectors of which are called Evaluation, Potency, and Activity (EPA). The EPA space is hypothesized to be a universal organizing principle of human socio-emotional experience, based on the discovery that these dimensions structure the semantic relations of linguistic concepts across languages and cultures (Osgood, May, and Miron 1975). They also emerged from statistical analyses of the co-occurrence of a large variety of physiological, facial, gestural, and cognitive features of emotional experience (Fontaine et al. 2007), and relate to the universal dimensionality of personality, and social cognition (Scholl 2013). EPA profiles of concepts can be measured with the semantic differential, a survey technique where respondents rate affective meanings of concepts on numerical scales. In general, within-cultural agreement about EPA meanings of social concepts is high even across subgroups of society, and cultural-average EPA from as little as a few dozen survey participants are extremely stable over extended periods of time (Heise 2010). Sociologists have gathered EPA ratings for a large number of concepts across different cultures (USA, Canada, Japan, Germany) by surveying thousands of people and compiled them into ACT lexicons or dictionaries which give average EPA ratings for words for which there is consensus (all ratings agree to a certain extent). For example, the EPA for the identity of *nurse* is [1.7,0.9,0.3], meaning that *nurses* are seen as quite good (E), a bit powerful (P), and a bit active (A)[1]. Comparatively a *patient* is seen as [0.9,-0.7,-1.1], less powerful and less active than a *nurse*. *Nurses* are expected to do positive and powerful things to a *patient* like *comfort* [1.5, 1.7, -0.6], rather than more negative and weak behaviours like *ignore* [-1.6, -0.8, -1.4].

Social events cause transient impressions of identities and behaviors that deviate from their corresponding fundamental sentiments (Heise 2007). ACT models this formation of impressions from events with a minimalist grammar of the form actor-behavior-object. Consider a *nurse* (actor) who *ignores* (behavior) a *patient* (object). Observers agree, and ACT predicts, that this *nurse* appears less nice (E), and less potent (P), than the cultural average of a *nurse*. The ACT predictions are based on a set of empirically derived non-linear impression formation equations (Heise 2007). The weighted sum of squared Euclidean distance between fundamentals and transients is called the *deflection* and is hypothesized to correspond to an aversive state of mind that humans seek to avoid or minimize (the *affect control principle*). The *nurse* who *ignores* a *patient* has a deflection of over 15 (very high), whereas if the *nurse comforts* the *patient*, the deflection is 1.5 (very low). The affect control principle allows ACT to compute deflection minimizing actions for agents using derivatives of the impression formation equations.

Identities can also be combined with modifiers using empirically derived non-linear impression formation equations (Heise 2007). That is, the identity for *nurse* as above can be combined with an affective modifier like *irritable* [-1.8,-0.7,0.8] to get a combination identity for an *irritable nurse* of [-1.0,0.3,0.5], considerably less nice and less powerful and a bit more active than a (non-irritable) *nurse*.

User interactions with online self-assessments

The goal was to simulate basic elements of the interactions of older adults with dementia (OAD) when using online self-assessments (online memory tests: OMT), grounding the simulations on the basic premiss that OAD will respond to OMT in an emotional way, according to the principles of ACT. A list of identities for OMT was therefore established

[1]EPA values range from -4.3 to 4.3 by convention. All ratings in this paper are taken from the Indiana 2002-2004 survey unless otherwise noted.

URL	Test name	identities based on	Identities	Modifiers
www.alzheimersreadingroom.com/2009/03/mini-cog-test-for-alzheimers-and.html	Mini-Cog	website creator's bio and photo	Son, **Caregiver**, CEO/executive, Expert, Writer	Smiling, **Understanding**, Caring
cft3.foodforthebrain.org/	Food for the Brain Cognitive Function Test	About us page, by line	Not-for-profit, clinic, **nutritionist**, charity, foundation, problem-solvers	**Scientific**, Charitable, Educational, Healthy
www.alzcast.org	Dementia Risk Assessment	About us for John Hopkins School of Medicine	Physicians, Scientists, Health Care System, **University**	**Reputable**, Rich, Promising, Elite
www.alzheimersprevention.org/alzheimers-info/memory-quiz	Memory Quiz (AP)	About Dr. Khalsa	**Doctor**, Pioneer, Writer, Founder, Sikh	Holistic, Natural, Reputable, Decorated, **Sikh**, Smiling
www.dailymail.co.uk/health/article-2095705/Alzheimers-questionnaire-Test-reveal-YOU-risk.html	Dementia ... or Simple Absent-Mindedness	About Banner Sun Health Research Institute	**Health care** system, Hospital services, Employer, Nonprofit	**Compassionate**, Professional, Caring, Available
www.memozor.com/memory-tests/test-for-alzheimer-online	quick test Alzheimer	Home page mission information	**Entertainer**, Game Player	**Generous**, Fun, Helpful
www.onmemory.ca/en/signs-symptoms/memory-test	Memory Test (OnMemory)	About us for the Alzheimer Society of Canada	Not-for-profit, **Charity**, Health Organization	Leading, Canadian, **Nearby**, Helpful, Transparent/Open
www.theweek.co.uk/health-science/61214/take-the-dementia-test-and-find-out-your-brain-age	How old is your brain?	Information contained in article about person (Dr. VincentFortanasce) who created the test	**Professor**, Academic, Researcher	**American**, Eminent

Table 1: Included online memory tests showing URLs and test names. The third column gives the information on which the assigned identities were ascribed. The most salient identitites and modifiers for each test are in the rightmost two columns

based on information about the site's creators or maintainers (e.g. from the "about us" sections). A list of identities for OAD was then established based on our previous qualitative work on identity in dementia. Finally, a simulation was carried out to predict whether or not the OAD would follow the advice of the OMT. A comparative analysis of the different OMT-OAD combinations can subsequently inform how to provide advice or to structure OMT according to the affective identities of end users.

OMT identities A set of eight online memory tests were considered. The tests were taken from an updated sample of a previous study of online self-assessments for dementia (Robillard et al. 2015). Identities and modifiers were assigned to each of these tests as follows. If a clear individual was visible or identified on the main page of the test, a set of identities and modifiers was assigned based on the characteristics of that individual. If no such individual was visible or identified, the individual(s) or organizations that created or hosted each test were traced by navigating within the domain of the test and identities and modifiers were assigned to those individuals/organizations instead. Table 1 shows the eight tests in our sample for the present study. The boldfaced identities and modifiers are those that were most salient. Three baseline identities were also included as follows (1) a generic *website* identity (with no modifier); (2) a *greedy salesman*; and (3) a *cultish guru*.[2]

OAD Identities A set of identities for OAD was assembled from two sources: 1) an interview study with older adults (König et al. 2017), and 2) an ongoing survey study of older adults with dementia.

Our recent qualitative work has been focussed on gaining a deeper understanding of identity in Alzheimer disease and dementia in the context of the ACT@HOME project (König et al. 2017; 2016). The long-term goal of ACT@HOME is to build cognitive assistive technologies that are emotionally aligned with OAD, as described in (Lin et al. 2014). The

[2]The last two baselines were based on the interpretations of the authors of Dr.Khalsa's `www.alzheimersprevention.org` page (see Table 1), on which he is promoting his "brain caps": supplements which contain natural lipids, oils and herbs (Ginkgo).

			EPA profiles		
type	identity	modifier	identity	modifier	combination
OMT	caregiver	understanding	[2.1 1.2 0.1]	[2.3 1.6 -0.2]	[2.4 1.1 -0.3]
OMT	nutritionist	scientific	[1.0 0.8 0.2]	[0.7 0.9 -0.2]	[0.6 0.6 -0.2]
OMT	university	reputable	[1.1 1.1 0.5]	[1.1 1.4 1.1]	[0.6 1.0 0.6]
OMT	doctor	Sikh	[1.9 0.7 0.1]	[-0.1 -0.1 -0.6]	[0.6 0.2 -0.6]
OMT	health_care	compassionate	[0.7 0.9 0.0]	[1.3 1.1 1.1]	[0.4 0.8 0.3]
OMT	entertainer	generous	[1.6 2.1 1.8]	[3.0 1.9 1.1]	[2.1 1.9 1.5]
OMT	charity	nearby	[1.4 1.2 1.2]	[1.0 0.3 -0.1]	[0.9 0.7 0.5]
OMT	professor	American	[1.6 1.6 0.3]	[0.7 1.1 0.8]	[0.7 1.2 0.3]
BLI	website		[1.3 0.9 0.9]		[1.3 0.9 0.9]
BLI	guru	cultish	[1.1,1.3,0.5]	[-0.6,0.02,0.2]	[-0.3,0.7,0.2]
BLI	salesman	greedy	[0.1,0.5,2.0]	[-1.4,0.4,0.5]	[-1.2,0.3,1.3]
OAD	senior	healthy	[1.1 0.6 0.9]	[3.2 2.4 0.9]	[1.9 1.0 0.8]
OAD	secretary	caring	[1.0 0.0 -0.2]	[2.3 1.7 0.4]	[1.4 0.4 -0.2]
OAD	husband	bipolar	[1.7 1.4 1.1]	[-0.7 -0.5 -0.2]	[0.0 0.5 0.3]
OAD	lawyer	irritable	[0.5 1.1 0.8]	[-1.8 -0.7 0.8]	[-1.3 0.3 0.6]
OAD	protector	german	[1.2 1.2 0.2]	[-0.5 -0.3 -0.2]	[-0.1 0.5 -0.3]
OAD	wife	athletic	[2.3 1.4 1.5]	[0.7 1.1 1.4]	[0.9 1.2 1.4]
OAD	slob	self centered	[-1.2 -1. -0.8]	[-0.9 0.1 0.1]	[-1.3 -0.8 -0.7]
OAD	homemaker	fortunate	[0.9 0.3 -0.1]	[2.3 1.9 0.9]	[1.3 0.7 0.1]
OAD	immigrant	proud	[0.5 0.4 0.7]	[2.1 2.1 1.5]	[0.8 0.8 0.9]
OAD	patient	sick	[1.9 1.2 -0.2]	[-1.5 -1.1 -0.7]	
		anxious		[-0.8 -0.3 0.9]	[-1.0 -0.2 -0.5]

Table 2: Learned EPA ratings for each identity and modifier for older adults with dementia (OAD), online memory tests (OMT) and baselines (BLI).

initial phase of the project has been a set of qualitative interviews with 12 OAD and their caregivers. The interviews covered life domains (e.g., family, origin, occupation), and about feelings related to intelligent cognitive assistants. Interviews were carried out with both caregivers and OADs. The caregiver interviews asked questions about a specific older adult with whom the caregiver was familiar. Interviews were transcribed and analyzed manually to extract a set of affective identities and modifiers (König et al. 2017). A subset of eight these identities and modifiers is shown in Table 2 (second and third columns, labeled OAD).

The second source of identities is a survey of OADs and their online health information-seeking habits (in progress). Data from this work show that anxiety is a key dividing factor in survey respondents. Therefore, a generic set of two identities as for *healthy senior* or *sick anxious patient* was added to the set.

Simulations As not all the identities and modifiers existed in the ACT lexicons, an EPA profile for each OAD and OMT identity label was computed using a supervised label propagation method (see below). The profiles were computed for the most salient identities and modifiers for OMT (shown in bold in Table 1, and repeated in Table 2), and for each OAD identity-modifier pair (shown in Table 2).

Based on the distributional linguistic hypothesis stating that words that occur in similar contexts tend to have a similar meaning (Harris 1981; Firth 1957), a support vector regression (SVR) model was trained on a pre-trained word embedding model called *word2vector* (Mikolov et al. 2013) and ACT lexicon (Francis and Heise 2006). Word2vector is a distributed representation of words based on their co-occurrence statistics obtained by training a recurrent neural network and skip-gram model on the Google News dataset, which has around 100 billion words (Mikolov et al. 2013).

The ACT lexicon was divided randomly into training (864 words) and testing (368 words), excluding words that do not exist in the word2vector model. A total of 256 words were excluded from both the training and testing sets (e.g., *public defender*, *talk down to*, and *skid row*). A grid search found the best scores for the parameters of an RBF kernel: $c = 10$, $\gamma = 0.0$. To evaluate the effectiveness of SVR in computing EPA score of the words that do not exist in the ACT lexicon, the computed EPA scores were compared to their original values in the testing set. The results show that the Kendall τ (Kendall 1938) correlation between the imputed EPA scores and their original values were equal to 0.624, 0.526, and 0.492 for E, P, and A respectively. The F1-score of comparing the binary value (0=negative, 1=positive) of the imputed EPA against their original value was equal to 0.83, 0.77, and 0.81 for E, P, and A respectively. The ranking correlation (Kendall τ) scores between the induced EPA scores and their true EPA values are equal to (τ=0.612, p-value<0.001), (τ= 0.536, p-value<0.001), and (τ= 0.496 ,p-value<0.001) for E, P, and A respectively.

The impression formation equations from (Heise 2007) were then used to compute a single EPA rating for each identity-modifier combination (see ACT introduction section), as shown in the last column of Table 2. Finally, ACT simulations for OAD identity and each OMT identity gave estimates of emotional alignment for each pair. A sequence of two simple actions was simulated: (1) OMT *advise* [2.1, 1.6, 1.0] OAD; (2) OAD *obey/disobey* ([0.5,-0.5, -1.0]/[-1.4, 0.1, -0.1]) OMT. The ACT simulations updated the transient impressions dynamically after each behavior, and the deflection was computed after the second behavior (the OAD response: d_o for obey and d_d for disobey.). Likelihood ratios for the OAD *obeying* the OMT advice versus *disobeying* are computed as e^{-d_o}/e^{-d_d}. A second set of behaviors was examined: *agree/disagree with* ([1.7,1.3,0.9]/[0.1,0.8,0.6], respectively).

Results

Automated quality analysis of online health information

Table 3 shows the classification of the 50 documents using the leave-one-out cross-validation method. There were 26 documents with complementarity $C = 1$, and 24 without. Of the 26 documents with $C = 1$, 21 were identified by the decision tree learned on the other 49 documents. This method yielded a high precision (1.0) classifier, with a recall of over .80 and a F1-score of .9. The first split of every decision tree learned was on the word *doctor*. In most cases, this was able to correctly identify all the positive cases (paragraphs with complementarity). Interestingly, not a single document that did not meet complementarity contained the word *doctor*, although two contained the word *doctors*, and four contained the word *physician*. The five $C = 1$ documents that were missed (classified as $C = 0$, false negatives) contained statements about consulting a *physician*, *GP*, *pharmacist* and *healthcare provider*. However, these words are also used extensively in other non-complementary contexts, while *doctor* is not.

actual	predicted C		
C	1	0	total
1	21	5	26
0	0	24	24
total	21	29	F: 0.9

Table 3: Cross-validation results on 50 web-pages for $C =$ complementarity. The precision is 1.0 and recall is 0.81 yielding an F1 score of 0.9.

Characterization of emotional alignment in online self-assessments

The deflections and likelihood ratios that resulted from the simulations of relationships between OAD and OMT are shown in Table 4 for each dyad of OAD-OMT identities. Considering the average deflections and ratios, some OMT are much more likely to be *obeyed* than *disobeyed*: those with the most positive (E) and poweful (P) identities (Mini-Cog and quick-test Alzheimer). The *greedy salesman* is the least likely to be *obeyed*, and the *cultish guru* is the second least likely for 8/10 OAD identities. Similarly, some OAD identities were found to be more likely to *obey*: those with more positive and powerful identities, as it is unlikely for a positive, powerful identity to perform a negative act such as *disobey*. For some identities, it is more likely that they will *disobey* advice. An *irritable lawyer*, for example, would only respond to the quick-test Alzheimer (generous entertainer), all other likelihood ratios being less than 1.0, meaning this identity would be more likely to disobey the advice of the OMTs. A *self-centered slob* would be unlikely to respond to the *scientific nutritionist*, the *Sikh doctor*, the *cultish guru* or the *greedy salesman*, but would respond to other OMT identities. This provides evidence for paying careful attention to how OMT are framed in terms of perceived identity in OAD, and calls OMT to be developed using a highly participatory co-design methodology in which perceived identities are carefully evaluated during OMT development.

Similar results were obtained with the *agree/disagree* behaviors. However, the overall deflections for the negative actions (disagree as compared to disobey) were less extreme. Further, a shift was observed in the actions of specific OAD

		OAD modifier identity										
Website	OTM modifier identity	healthy senior	caring secretary	bipolar husband	irritable lawyer	german protector	athletic wife	self centered slob	fortunate homemaker	proud immigrant	sick anxious patient	Average
Mini-Cog	understanding caregiver	3.6/10.8 **1337.9**	2.5/9.0 **621.7**	3.6/6.0 **10.4**	6.3/6.0 **0.8**	3.5/6.0 **11.8**	4.5/8.3 **43.6**	5.2/6.2 **2.7**	2.8/8.5 **277.2**	3.5/7.3 **41.8**	4.8/5.8 **2.8**	4.1/7.4 **235.0**
Food for the Brain	scientific nutritionist	2.9/7.2 **74.7**	1.5/5.4 **51.0**	1.4/2.1 **2.1**	2.9/1.6 **0.3**	1.2/2.1 **2.5**	2.9/4.6 **5.4**	2.0/1.9 **0.9**	1.6/4.9 **26.1**	1.9/3.6 **5.7**	1.7/1.6 **0.9**	2.0/3.5 **17.0**
Dementia Risk Amt.	reputable university	2.9/7.3 **84.5**	1.5/5.6 **60.2**	1.4/2.5 **2.8**	3.0/2.0 **0.4**	1.3/2.5 **3.3**	3.0/4.9 **6.6**	2.1/2.3 **1.3**	1.6/5.1 **31.2**	1.9/3.9 **7.1**	1.8/2.0 **1.2**	2.0/3.8 **19.9**
Memory Quiz	Sikh doctor	3.1/7.2 **63.4**	1.6/5.3 **41.4**	1.5/1.9 **1.6**	3.0/1.4 **0.2**	1.3/2.0 **1.9**	3.1/4.5 **4.2**	2.1/1.7 **0.7**	1.8/4.8 **20.9**	2.0/3.5 **4.4**	1.8/1.4 **0.6**	2.1/3.4 **13.9**
Dementia/ Simple AM	compassionate health-care	2.9/7.0 **63.0**	1.4/5.2 **45.4**	1.3/2.0 **2.2**	2.8/1.6 **0.3**	1.1/2.1 **2.6**	2.9/4.5 **5.0**	1.9/1.9 **1.0**	1.6/4.7 **23.7**	1.8/3.5 **5.5**	1.6/1.6 **1.0**	1.9/3.4 **15.0**
quick test Alzheimer	generous entertainer	4.1/11.2 **1180.4**	3.0/9.5 **643.1**	4.0/6.7 **15.5**	6.5/6.8 **1.4**	3.9/6.8 **17.9**	4.9/8.9 **52.7**	5.5/7.0 **4.8**	3.3/9.0 **302.6**	4.0/7.9 **52.9**	5.0/6.6 **5.0**	4.4/8.1 **227.6**
Memory Test	nearby charity	2.6/7.3 **112.6**	1.2/5.5 **72.2**	1.3/2.3 **2.6**	3.1/1.9 **0.3**	1.2/2.3 **3.1**	2.8/4.8 **7.1**	2.1/2.2 **1.1**	1.4/5.0 **36.2**	1.7/3.8 **7.5**	1.8/1.9 **1.0**	1.9/3.7 **24.4**
How old is your brain?	American professor	3.0/7.8 **115.7**	1.6/6.0 **81.4**	1.6/3.0 **3.7**	3.3/2.6 **0.5**	1.5/3.0 **4.3**	3.1/5.3 **8.8**	2.4/2.9 **1.6**	1.8/5.5 **42.0**	2.1/4.3 **9.5**	2.1/2.5 **1.6**	2.3/4.3 **26.9**
baseline	website	2.7/8.1 **211.0**	1.4/6.2 **126.3**	1.8/3.1 **3.9**	3.7/2.9 **0.4**	1.6/3.1 **4.5**	3.1/5.5 **11.6**	2.8/3.1 **1.4**	1.6/5.7 **61.7**	2.0/4.5 **11.9**	2.4/2.8 **1.4**	2.3/4.5 **43.4**
baseline	cultish guru	3.5/6.7 **26.7**	1.9/5.0 **22.1**	1.4/1.8 **1.5**	2.6/1.2 **0.2**	1.2/1.8 **1.8**	3.2/4.3 **2.8**	1.8/1.6 **0.8**	2.0/4.5 **12.1**	2.1/3.3 **3.2**	1.5/1.3 **0.8**	2.1/3.2 **7.2**
baseline	greedy salesman	5.2/7.3 **8.1**	3.5/5.6 **7.9**	2.7/2.5 **0.8**	3.6/1.9 **0.2**	2.5/2.5 **1.0**	4.7/4.9 **1.2**	2.8/2.2 **0.6**	3.6/5.1 **4.6**	3.6/3.9 **1.4**	2.6/1.9 **0.5**	3.5/3.8 **2.6**
	Average	3.3/8.0 **298.0**	1.9/6.2 **161.2**	2.0/3.1 **4.3**	3.7/2.7 **0.5**	1.9/3.1 **5.0**	3.5/5.5 **13.6**	2.8/3.0 **1.5**	2.1/5.7 **76.2**	2.4/4.5 **13.7**	2.5/2.7 **1.5**	

Table 4: Matrix of deflection scores for each person identity (row) to "obey/disobey" each test identity (column). The number in boldface shows the likelihood ratio, giving how much more likely this person (row) is to obey this test (column) than to disobey, after an initial action of "advise" by the website.

	caring secretary	irritable lawyer	self-centered slob	sick anxious patient
agree	2.7	7.0	8.0	6.7
disagree	5.0	5.8	6.5	5.7
ratio	**10.2**	**0.3**	**0.2**	**0.4**
obey	3.0	6.5	5.5	5.0
disobey	9.5	6.8	7.0	6.6
ratio	**643.1**	**1.4**	**4.8**	**5.0**

Table 5: Comparison of different behaviours showing the deflections and likelihood ratio caused when interacting with the "quick test Alzheimer" OMT.

identities. For example, an *irritable lawyer*, *self-centered slob* and *sick anxious patient* will be all more likely to *obey* a *generous entertainer* (identity for the "quick test Alzheimer" OMT) than to disobey, but are more likely to *disagree with* than *agree with* the same test (see Table 5). Thus, the framing of the advice being given will be important for determining the level of engagement of OAD with the test results and recommendations.

Discussion

Results presented here demonstrate how artificial intelligence can be applied to the evaluation of the online health environment. Specifically, it was shown that 1) quality evaluation of static, online health resources can be carried out through automation of the QUEST tool (Robillard and Feng 2016); 2) affective analysis of interactive health resources can be conducted by applying the principles of affect control theory (Heise 2007) to interactions between the perceived affective identities of self-assessment providers and consumers.

The first step in the automation of the QUEST tool was learning a decision tree classifier for the complementarity criterion. This method yielded a reliable automation process for the classification, with a recall of over .80 and a F1-score of .9. These results confirm the possibility of predicting this simple criteria fairly robustly with a completely data-driven model. The next steps in this effort will be to integrate more prior knowledge into the classifiers (e.g. looking for combinations of words rather than bag-of-words), and to automate the other criteria. While the complexity of the remaining criteria in the QUEST tool varies widely, from the identification of authors and dates to more elaborate evaluations of tone and attribution, results presented here confirm that the nature of the criteria in QUEST is amenable to automation. Providing automated quality evaluations of online health resources broadly construed constitutes an innovative departure from previous attempts at controlling health information quality, such as quality seals or web-based resource hubs. By providing a tool to evaluate quality regardless of the source, QUEST automation is a more promising solution for the challenges associated with health-information seeking-behaviors that involve consulting a large variety of online sources.

The affective analysis of various combinations of identities for OADs and OMTs have shown that there is a variability in affective interpretations across different dyads. However, differences in affective interpretations contribute only one aspect of a person's decision-making process. Other aspects stem from cognitive reasoning about the advice provided in the OMT, such as background knowledge and the congruence of the advice with previously held beliefs about dementia as provided through social interactions (e.g., with family, friends, physicians). Nevertheless, a person's final decision will be biased by their affective interpretations of online health information and recommendations. According to Affect Control Theory, the strength of this bias will be proportional to the deflections caused. For example, a *healthy senior's* decision will be more heavily biased towards *obeying* an *understanding caregiver* than it would be towards a *greedy salesman*. The prediction of the theory is that this person will be more likely to cognitively evaluate the *greedy salesmean's* than the *understanding caregiver's* advice. The reaction to the *understanding caregiver* will be more automatic, and will call upon less cognitive resources. These findings are especially relevant in the context of the aging population, as cognitive impairment in dementia may shift the balance between affective and cognitive contributions to health decision-making.

The identities used in the OMT analysis presented here are only a first step in the direction of gaining an understanding of how emotional alignment plays into the adoption and rejection of health recommendations. Future work is needed to rapidly and reliably assign identities to various forms of media and efforts in this area will involve, for example, automated assessments of identities based on sentiment analysis of text. Similarly, the process of establishing OAD identities will also require streamlining and methods are being developed to achieve this goal through validated survey instruments. Reliable assignment of identities for both OAD and OMT will allow for the assessment of the alignment criteria as a predictive tool of online health information usage and will inform the development and production of virtual assistants to assist OADs in processing and interpreting both static and interactive online resources. Our recent work on a probabilistic and decision theoretic generalization of the ACT model may be useful in dealing with the multiple identities that can be interpretations of online information, and of the multiple identities held by older adults with dementia. (Hoey, Schröder, and Alhothali 2016; Schröder, Hoey, and Rogers 2016; König et al. 2017).

Conclusion

Unregulated online resources for dementia receive millions of unique visitors each month. It is critical to assist older adults with dementia (OAD) in avoiding negative health outcomes, financial harm, and inappropriate use of health care services that result from consulting these resources. Taken together, the intelligent and emotionally-aligned processes proposed in this paper lay the groundwork for the development of innovative tools to assist OADs and their caregivers safely navigate the online environment. Building on our earlier work that established validated metrics for evaluations of quality, ethics and access of online resources about dementia (Robillard and Feng 2016) as well as our previous work on emotionally aware intelligent virtual assistants (Malhotra et al. 2016), these tools will provide guidance to users in a way that is aligned with end-users identities and that respects their values. Further, results presented here and future work in this area will generate new knowledge on the integration of identity theories into health technologies to promote adoption, adherence to ethical norms, and legitimacy.

Acknowledgments

We thank Tanya Feng for annotation work. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and by AGE-WELL NCE Inc., a member of the Networks of Centres of Excellence program. A. Alhothali is supported by a grant from King Abdulaziz University of Saudi Arabia.

References

Benigeri, M., and Pluye, P. 2003. Shortcomings of health information on the internet. *Health Promotion International* 18(4):381–386.

Blank, G., and Lutz, C. 2016. Benefits and harms from internet use: A differentiated analysis of great britain. *New Media Society*.

Breiman, L.; Friedman, J.; Olshen, R.; and Stone, C. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth.

Charnock, D.; Shepperd, S.; Needham, G.; and Gann, R. 1999. Discern: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health* 53(2):105111.

Chumber, S.; Huber, J.; and Ghezzi, P. 2015. A methodology to analyze the quality of health information on the internet the example of diabetic neuropathy. *The Diabetes Educator* 41(1):95–105.

Corner, L., and Bond, J. 2004. Being at risk of dementia: Fears and anxieties of older adults. *J Aging Stud* 18(2):14355.

Eysenbach, G.; Powell, J.; Kuss, O.; and Sa, E.-R. 2002. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *Jama* 287(20):2691–2700.

Firth, J. R. 1957. *A synopsis of linguistic theory, 1930-1955*. Oxford: Blackwell.

Fontaine, J. R. J.; Scherer, K. R.; Roesch, E. B.; and Ellsworth, P. C. 2007. The world of emotions is not two-dimensional. *Psychological Science* 18:1050 – 1057.

Francis, C., and Heise, D. R. 2006. Mean affective ratings of 1,500 concepts by indiana university undergraduates in 2002-3. *Data in Computer Program Interact*.

Giles, D., and Newbold, J. 2011. Self- and other-diagnosis in user-led mental health online communities. *Qual Health Res* 21(3):41928.

Harris, Z. S. 1981. *Distributional structure*. Springer.

Hartin, P. J.; Nugent, C. D.; McClean, S. I.; Cleland, I.; Tschanz, J.; Clark, C.; and Norton, M. 2016. The empowering role of mobile apps in behavior change interventions: The gray matters randomized controlled trial. *JMIR mHealth and uHealth* 4(3:e93).

Heise, D. R. 2007. *Expressive Order: Confirming Sentiments in Social Actions*. Springer.

Heise, D. R. 2010. *Surveying Cultures: Discovering Shared Conceptions and Sentiments*. Wiley.

Hoey, J.; Schröder, T.; and Alhothali, A. 2016. Affect control processes: Intelligent affective interaction using a partially observable Markov decision process. *Artificial Intelligence* 230:134–172.

Hsu, W.-C., and Bath, P. 2008. Development of a patient-oriented tool for evaluating the quality of breast cancer information on the internet. *Stud Health Technol Inform* 136:297302.

Jones, J. 1999. Development of a self-assessment method for patients to evaluate health information on the internet. *Proc AMIA Symp* 540544.

Kendall, M. G. 1938. A new measure of rank correlation. *Biometrika* 30(1/2):81–93.

König, A.; Francis, L. E.; Malhotra, A.; and Hoey, J. 2016. Defining affective identities in elderly nursing home residents for the design of an emotionally intelligent cognitive assistant. In *Proc. International Conference on Pervasive Computing Technologies for Healthcare*, 206–210.

König, A.; Francis, L.; Joshi, J.; Robillard, J.; and Hoey, J. 2017. Qualitative study of affective identities in dementia patients for the design of cognitive assistive technologies. *Journal of Rehabilitation and Assistive Technologies Engineering*. in press.

Libert, T. 2015. Privacy implications of health information seeking on the web. *Commun. ACM* 58(3):6877.

Lin, L.; Czarnuch, S.; Malhotra, A.; Yu, L.; Schröder, T.; and Hoey, J. 2014. Affectively aligned cognitive assistance using bayesian affect control theory. In *Proc. of International Workconference on Ambient Assisted Living (IWAAL)*, 279–287. Belfast, UK: Springer.

Lovett, K.; Mackey, T.; and Liang, B. 2012. Evaluating the evidence: direct-to-consumer screening tests advertised online. *J Med Screen* 19(3):14153.

MacKinnnon, N. J., and Robinson, D. T. 2014. 25 years of research in affect control theory. *Advances in Group Processing* 31.

Malhotra, A.; Hoey, J.; König, A.; and van Vuuren, S. 2016. A study of elderly people's emotional understanding of prompts given by virtual humans. In *Proc. International Conference on Pervasive Computing Technologies for Healthcare*, 13–16.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.

Moult, B.; Franck, L.; and Brady, H. 2004. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health care information. *Health Expect Int J Public Particip Health Care Health Policy* 7(2):165175.

Norman, C.; McIntosh, S.; Selby, P.; and Eysenbach, G. 2008. Web-assisted tobacco interventions: empowering change in the global fight for the publics (e)health. *J Med Internet Res* 10(5:e48).

Osgood, C. E.; May, W. H.; and Miron, M. S. 1975. *Cross-Cultural Universals of Affective Meaning*. University of Illinois Press.

Reeves, B., and Nass, C. 1996. *The media equation*. Cambridge University Press.

Robillard, J. M., and Feng, T. L. 2016. Health advice in a digital world: Quality and content of online information about the prevention of alzheimers disease. *Journal of Alzheimer's Disease* preprint(preprint).

Robillard, J. M.; Illes, J.; Arcand, M.; Beattie, B. L.; Hayden, S.; Lawrence, P.; McGrenere, J.; Reiner, P. B.; Wittenberg, D.; and Jacova, C. 2015. Scientific and ethical features of english-language online tests for alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 1(3):281 – 288.

Ryan, A., and Wilson, S. 2008. Internet healthcare: do self-diagnosis sites do more harm than good? *Expert Opinion on Drug Safety* 7(3):2279.

Sandvik, H. 1999. Health information and interaction on the internet: a survey of female urinary incontinence. *BMJ* 319(7201):29–32.

Scholl, W. 2013. The socio-emotional basis of human interaction and communication: How we construct our social world. *Social Science Information* 52:3 – 33.

Schröder, T.; Hoey, J.; and Rogers, K. B. 2016. Modeling dynamic identities and uncertainty in social interactions: Bayesian affect control theory. *American Sociological Review* 81(4).

Seidman, J.; Steinwachs, D.; and Rubin, H. 2003. Design and testing of a tool for evaluating the quality of diabetes consumer-information web sites. *J Med Internet Res* 5(4:e30).

Shank, D. B., and Lulham, R. 2016. Products as affective modifiers of identities. *Sociological Perspectives*.

Shank, D. B. 2010. An affect control theory of technology. *Current Research in Social Psychology* 15(10):1–13.

Silberg, W. M.; Lundberg, G. D.; and Musacchio, R. A. 1997. Assessing, controlling, and assuring the quality of medical information on the internet: Caveant lector et vieworlet the reader and viewer beware. *Jama* 277(15):1244–1245.

Wolfe, S. M. 2002. Direct-to-consumer advertising education or emotion promotion? *New England Journal of Medicine* 346(7):524–526.

Zhang, Y.; Sun, Y.; and Xie, B. 2015. Quality of health information for consumers on the web: A systematic review of indicators, criteria, tools, and evaluation results. *J Assoc Inf Sci Technol* 66(10):20712084.