Demystifying AI in a call center environment

AI and Machine Learning

Demystifying AI in a call center environment, using predictive analytics and affective computing to drive performance

NAME(S) OF AUTHOR: KYLE BARKER & ABID SHAH

Cryptek Labs

E-mail: info@crypteklabs.com

Abstract

Affective computing technologies are designed to sense and respond based on human emotions (Brigham, 2017). While so much personal information is digitally quantified now — including what we read, what we buy, and even our own health—there remains a disconnect in most Artificial Intelligence (AI) applications between data and human emotion (Mok, 2015). However, technology has evolved to a point where it is capable of processing vast amounts of data quickly, to assess not only what an individual is saying but also their emotional state. Systems can offer the appropriate response based on what is felt, bringing benefits to working environments such as call centers, as well as other critical areas, like fraud detection and customer vulnerability.

With the exponential growth in voice activated Internet Of Things (IoT) platforms and desire for hyper-personalized customer experiences, some research has cited the limitations of an affective approach (Landowska et al, 2016). Yet human emotion has a significant role to play for businesses looking to remain competitive within a crowded digital market.

This paper provides an overview of affective computing along with its benefits and limitations, as well as case studies of how, and where, it is currently being used.

KEYWORDS: Affective computing, predictive analytics, emotion, human interaction, sentiment analysis, natural language processing

Call
Toll free: +1 877 239 3153

Email
info@crypteklabs.com
crypteklabs.com

Demystifying AI in a call center environment

AI and Machine Learning

1. Introduction

Up until the mid-1990s, interface and technology design were primarily focused on the cognitive constraints of the user (Calvo et al, 2013). Affective experiences such as emotions, moods and feelings were not really on the radar until the realization that humans are more than just cognitive machines and their continual interplay with emotions is the basis of information processing in humans. This recognition led to what some are calling the "affective revolution" and the rise of the field of Affective Computing (Picard, 1997).

Affective computing, which is sometimes known as artificial emotional intelligence or emotion AI (Kaliouby, 2017), is the study of systems and devices that can recognize, interpret, process and simulate human affects. The machine should be able to interpret the emotional state of a human and give an appropriate response to the situation.

Much progress has been made in the field over the last decade. This paper examines those achievements, as well as the benefits and potential limitations of the technology in its more scientific-use cases.

2. Applications of Affective Computing

Programs are being developed that can use facial expressions and micro-expressions, posture, gestures, tone of voice, speech, and even the rhythm or force of keystrokes as well as the temperature of one's hands, to register changes in a user's emotional state (Marr, 2016). Cameras and other types of sensors take the input data and process it through deep learning algorithms, which are able to determine an emotional state and then make a decision accordingly. In reality, the applications are limitless, as virtually every decision made by humans involves an emotion of some description.

EMOTIONAL SPEECH

Speech produced in different states such as anger, fear, or joy can be interpreted by an affective computing system and translated into an action. This is done by using various speech features, which are discussed in more detail later in this paper.

Demystifying AI in a call center environment

AI and Machine Learning

FACIAL DETECTION

Emotion AI technology can detect and process facial expressions using methods like optical flow, neural networks, and hidden Markov models (a statistical model often used in machine learning or data mining) . While these models are not the focus of this paper, it should be noted that when combined with speech, they are able to provide a better overall view of emotion in a face-to-face context.

BODY GESTURES

Affective computing systems can analyze the way we interact with, or touch, various objects as an indicator of emotion. Additionally, common methods such as detecting the position of one's arms or shoulders in a certain situation that can represent anxiety, are used for comprehensive profiling.

PHYSIOLOGICAL MONITORING

This involves monitoring and analyzing signs such as pulse or heart rate to recognize emotion.

While there are several areas that could benefit from emotion AI technology, some that are already making use of it include: e-learning, to detect when a user is having difficulty and offer additional guidance; e-therapy, to deliver health services online like counselling; and advertising, to review how users respond to different promotions or marketing. Beyond this, the technology can even be used to help people on the Autism spectrum interact with others, which is a major real-world impact (Picard, 2009).

3. Affective Computing in Call Centres

The traditional sales and service customer call center is an ideal application for emotional speech analytics and artificial intelligence. In these environments, agents handle calls continuously. By understanding the emotional state of the caller, they will not only be able to handle the call better, but also guarantee a more effective customer experience with the potential for reduced costs and increased income.

One of the key skills in a call center is the ability to understand emotional intelligence. This reflects how well a call center employee can empathize with,

communicate with, and influence the caller. The benefits lie in fostering a sense of connection (often known as building a rapport) with customers, regulating personal feelings to deal with hostile communication, and ensuring that every single caller receives the correct type of response.

Emotionally and socially sensitive AI can guide call center workers to better performance. As customers speak, the software can analyze the speech and provide real-time coaching of the dialogue to create better interactions. This is vitally important for employees who, perhaps, don't have strong skills in emotional intelligence. For example, the software could recommend that they speak more slowly, or help them recognize when they are interrupting the customer at the wrong time (Knight, 2017).

Call center software automatically assesses the dynamics of a conversation, as it has been trained to recognize pertinent characteristics. This is accomplished through machine-learning algorithms that absorb copious amounts of training data to learn from experience. Every new call that comes into the call center is an opportunity for learning and the technology can ingest this over a continuous incremental cycle.

As the technology develops, the customer experience will continue to evolve, creating new and customized journeys for everyone.

There is a two-fold benefit to emotion AI in that it can also provide benefits to the agents (Feast, 2016). Call center workers tend to take medical leave at around three times the average rate of other disciplines and have a churn rate of up to 26% annually. The profession can be stressful as most employees are taking back-to-back calls, much of which can be stressful depending on the nature of the industry (e.g. customers often phone to complain that something isn't right). Emotion AI could be used to analyze all different types of speech, and is hugely valuable in ensuring the wellbeing of staff as much customers.

4. Algorithms used in speech emotion detection

Detection and classification are two typical tasks for machine learning algorithms and more importantly, deep learning neural networks (Bonarik, 2019).

Call
Toll free: +1 877 239 3153

Email
info@crypteklabs.com
crypteklabs.com

Demystifying AI in a call center environment

AI and Machine Learning

Algorithms will learn from training data, such as a history of conversations and images, to take examples of what different emotions look like and how they can sometimes work simultaneously. This means that when a new section of speech is passed into the algorithm, it can use knowledge from the training set to help determine what the emotion was being displayed by the customer. Previous emotions have been classified or labelled in order to train the models appropriately.

Deep learning neural networks associate a particular set of emotion cues, coming from different channels with a particular set of emotions or cognitive states (Pilev, 2018). Multi-modal algorithms, those using both speech and facial data, have been more effective than uni-modal scripts as the channels can collaborate and verify the information received with one another.

Some of the key technical algorithms used would include:

- *LDC*—(Linguistic Data Consortium) classification of the emotion is based directly on which group it is associated to.
- *kNN*—sclassification of the emotion is based on the nearest result. If the algorithm cannot find an exact match, it will find the closest match, or nearest neighbor, as this is commonly referred to.
- *Decision tree*—a series of rules or paths work out which emotion the speech is classified into. The branches of the tree represent subsequent features.
- *HMMs*—This uses what is known as a Markov model to work out the probability of different emotional states and is one of the most common methods in speech detection.

Other machine learning methods are used in emotion AI applications but those named above tend to be the most common.

5. Case studies and results

Applications of affective computing and emotion AI technology have seen varying degrees of success. Below we outline several:

Call
Toll free: +1 877 239 3153

Email
info@crypteklabs.com
crypteklabs.com

Demystifying AI in a call center environment

AI and Machine Learning

www.crypteklabs.com

EMOTIONAL INTELLIGENCE AT HUMANA

Health insurance provider Humana has seen a 28% improvement in customer satisfaction scores since deploying their emotion AI technology. They achieved this through rapid experimentation and testing of hypotheses (Morgan, 2016). This process involves defining a problem (e.g. customers are leaving negative feedback), starting to understand what it is, and then developing a machine learning algorithm capable of resolving the situation. Humana has developed their own metric for calculating emotion as part of their project. They found a direct correlation between this and other core performance indicators.

MEDICAL EMERGENCY CALL CENTRE IN FRANCE

Research at a medical center in France have shown results to be at around 82% accuracy for detecting the correct emotions (Vidrascu et al, 2005). This was in the main differentiating between positive and negative emotions using paralinguistic cues without taking into account the conflicts between blended emotions.

EMOTION AI AT DISNEY

Disney has created algorithms to determine how much audiences like its movies. The complex algorithm takes in facial expressions and even has the ability to predict upcoming emotions (Gilliland, 2018).

The algorithm was tested during the Disney releases of ‘The Jungle Book’ and ‘Star Wars: The Force Awakens’ with a few minutes of tracking facial behaviour being enough to predict whether viewers were going to laugh or smile at specific moments during the movie. The testing was able to generate 16 million data points from 3,179 viewers which is far more than any human would be capable of, especially in real-time, resulting in a far more accurate response.

INTELLIGENT VOICE AND AUDIO RECOGNITION

Intelligent Voice, an audio recognition software that performs emotion AI, has started working with insurers and banks in the US and UK to detect potential fraud and conduct credit-worthiness assessments, all from the sound of the speaker’s voice.

Demystifying AI in a call center environment

AI and Machine Learning

The technology can tell if people are making up a story or being genuine from their tone of voice and could represent a multi-billion dollar opportunity for the insurance industry.

6. Speech descriptors

Some studies have suggested that there could be as many as 200 distinct features of speech used within real-time detection, which makes the tracking of emotion, especially in real-time, a very complex process. It is important that any technology used is able to ascertain which of those are redundant and optimize the results of the use case. These are often grouped together into three broader umbrella terms (Steidl, 2011):

1. *Frequency characteristics*—this can include aspects like the shape of accents, the level of pitch, and slope of contours that the speaker uses.
2. *Time-related features*—the speed at which the speaker is talking.
3. *Voice quality parameters and energy descriptors*—this will include features like breathlessness, pauses and loudness of the speaker.

Bondale & Sreebuvas (2012) refer to these markers as 'emotiphons,' which are short, lexical expressions in conversational speech conveying emotions by modifying the prosody of the utterance. It is thought that these are an unintentional kind of emotion and bring out a better result in terms of understanding the speaker, as opposed to more deliberate speech bursts (Schroeder, 2003).

The range of emotions covered by emotiphons are far more than those used within standard literature and databases. For example, the common emotions are anger, fear, joy and sadness, which are mainly recognized at an acoustic level. The concept of emotiphons is to take these to the next level by distinguishing the nuances of the emotional types, like affection, pain, sympathy, or disbelief. All of these are classified into sub-categories and groups for analysis and processing purposes.

Research in Darwinian theory indicates that all humans have an innate sense of emotions. This is supported by studies such as Hozjan & Kacic (2006) and Black & Yacoob (1995). However, it is the way that these emotions are expressed that is

Demystifying AI in a call center environment

AI and Machine Learning

important within the emotiphon theory, and this is where algorithms tend to slip up in their analysis of emotion AI.

Models such as that developed by Scherer & Mehu (2012), seen below, have proposed complex models for the development of emotion AI understanding in humans and these can be translated into the initial algorithms used by technology. Figure 1 shows the Tripartite model adapted from their research.

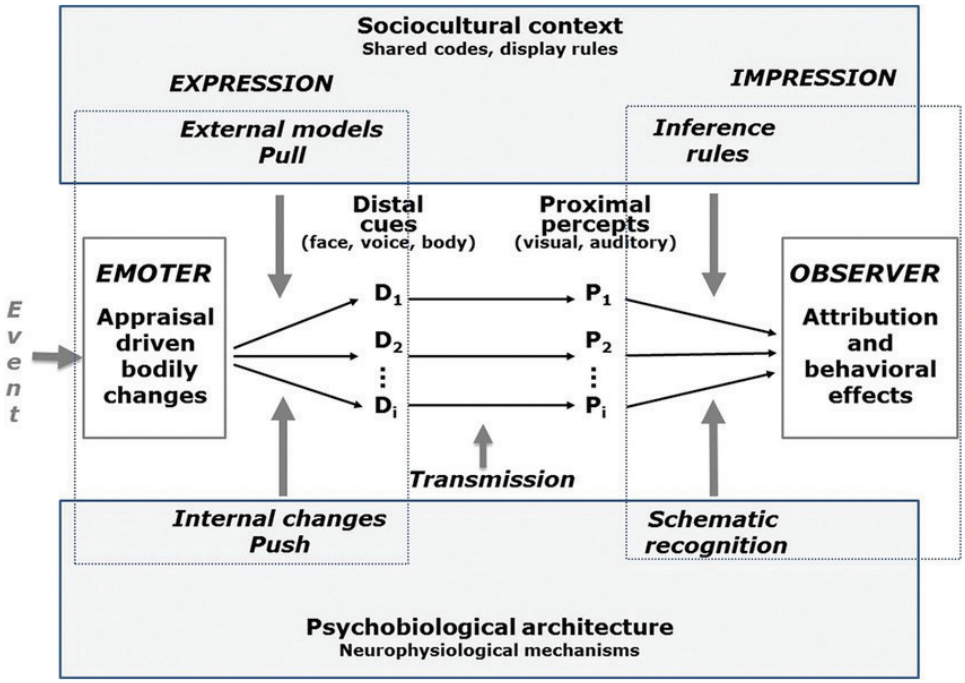

Figure 1 – The tripartite emotion and expression model

The concept of this model is that the communication process is represented by four elements (emoter/sender, distal cues, proximal concepts, and observer) and three phases (externalization, driven by external models and internal changes; transmission cue utilisation, driven by inference rules; and schematic recognition). Without going into technical detail about the underlying algorithms, if this model is applied to emotion AI, the internal state of the speaker is encoded by what

are known as “distal vocal cues” and the listener perceives the utterances and extracts several cues (e.g. voice quality ratings). It is these perceptions that are used to gather an internal state of the user based upon a set of rules.

Algorithms for affective computing must be structured in such a way that allows for the machine to recognize emotional cue within speech.

7. Hyper Connected Smart Devices

Everyday lives are becoming guided by smart technology and Internet of Things (IoT) devices as humans strive to stay connected. An IoT device is defined as anything within an internet connection such as a Smartphone, Amazon Alexa, wearables like Fitbit, or machinery in the manufacturing industry.

It’s expected that there will be more than 64 billion IoT devices worldwide by 2025 (Petrov, 2019) in a global market worth $1.7 trillion. What this means for affective computing is an ever-growing need to interpret emotions, as humans look for technology to know them just as well as they know themselves.

Koulibaly (2017) cites the Amazon Alexa, which has an AI model based around voice commands and is supposed to provide autonomous, conversational and relational responses. However, the issue with these devices at the moment is that they appear to be devoid of emotion and cannot tell if we are happy, angry, or depressed when stating a command.

Over the next three to five years, it is expected that the technologies surrounding us will become emotionally aware. As with any machine learning application, they learn from experience. We are now at a point where there is potentially enough big data and processing power to truly propel affective computing into the 21st century. Business insight from Gartner predicts that by 2022, 10% of personal devices will have emotion AI capabilities, up from less than 1% in 2018 (Forsdick, 2019).

8. Technical Limitations of Affective Computing

Existing affective computing technology has some limitations. First of all, we live

Call
Toll free: +1 877 239 3153

Email
info@crypteklabs.com
crypteklabs.com

in a diverse world and there will be cultural and dialect-based discrepancies. Consider the various accents across different US states or UK counties alone, and it is quite clear that there is a need to pick up on different language intonations and contexts. This will always be a challenge when developing any sort of emotion AI solution. It requires detailed mapping and algorithms. Machines will be able to learn about discrepancies over time but it might take some years for it to be entirely accurate.

In facial recognition, if we look at social media for instance, many facial expressions tend to be posed and therefore aren't natural. This would be challening for an algorithm to learn from, given that there isn't any real consistency. It can therefore make using data from such sources null and void, until that can be accounted for. There may be a need to exclude "selfies" from analysis, for example, and only use images which appear to be natural.

Beyond the technological limitations, as humans become more aware of emotion AI, there may be a propensity for them to act differently. For example, if callers knew that their speech may be used to assess fradulent behaviour, would they attempt to change how they speak to throw off any algorithm being used? The technology will need to continually evolve in order to determine whether speech changes are genuine or falsified.

Although the current technical limitations are important, it is arguably more apt to note the ethical implications of affective computing.

9. Ethical considerations of Affective Computing

Some recent television shows that focus on artfificial intelligence and technology, such as the cult series "Black Mirror," expose darker scenarios surrounding applications of affective computing. While these series are portrayed to be futuristic, given the power of processing from cloud and quantum computing, they may not be quite as distant anymore (Cooney et al, 2018).

It is thought that emotion AI has the capability to blur lines between what should be private and public information, with the potential for both physical and psychological harm. It is often the case that people are not even aware of

Call
Toll free: +1 877 239 3153

Email
info@crypteklabs.com
crypteklabs.com

Demystifying AI in a call center environment

AI and Machine Learning

their own emotions, let alone the fact that they are being analyzed by a robot or machine. While this does bring some benefit when trying to detect dangerous situations, there could be several pitfalls:

- *Weaknesses in human nature:* Referring back to the earlier example of “Black Mirror” (Black Mirror, 2017) , in one eposode a mother tracks the emotions of her daughter via a tablet in order to protect her from everything dangerous in the world. The manipulation of relationships and events ultimately leads to violence and causes more harm than good.

- *Psychological harm:* Revealing emotions through speech, facial recognition and body language could reveal things that we do not want to expose about ourselves. For example, it could show signs of depression or anxiety that we wanted to keep to ourselves rather than having an employer or insurer find out about it through technology.

- *Physical harm:* Some theorists (Bonn, 2017) have suggested that positive emotions can lead to anger and violence. Bullies might use emotion detection as a method of picking up on the fear of others in extreme circumstances.

- *Miscommunication:* At the moment, affective computing systems seem to act on the fact that people can only feel one emotion at any given time. As most of us know, emotions are far more complex than just dealing with one thing at a time. It would be very easy for machines to add their own bias to situations and believe something that isn’t true. Something that seems quite innocuous, like nodding, can have a number of different interpretations (Poggi et al, 2010).

- *Dehumanising emotions:* If emotions are being used in everyday behavioural analytics, there is a possibility that they would get completely dehumanised and just be seen as “data” or “information.” It would be bad for society of machines are not capable keeping some semblance of emotional reality.

Call
Toll free: +1 877 239 3153

Email
info@crypteklabs.com
crypteklabs.com

Demystifying AI in a call center environment

AI and Machine Learning

While there are various ethical considerations with affective computing technology, there are strategies that can be put in place to at least minimize, if not completely negate, them. Just like any other piece of data that we share, humans can be asked for consent before emotion tracking to ensure everything is in line with regulations such as General Data Protection Regulation (GDPR), which protects consumer rights (Burgess, 2019).

The technology needs to have the ability to enable and disable emotion detection if it is to avoid the common pitfalls. As algorithms become more advanced and go beyond emotions (e.g. learning the intention of a behaviour as well as simply interpreting it), the technology will be able to evolve simultaneously.

Regulation may be the ultimate answer to ensure emotion AI is used in the proper way but it would be wrong to recommend that alternative now within an indsustry that is still growing exponentially.

10. Conclusion

With rapid advancement in technology, big data, artificial intelligence and cloud computing, the understanding and predicting of humans emotions is no longer a fantasy. Speech, facial, and movement analysis algorithms have grown powerful enough to be able to analyze all features of interactions in granular detail, creating a number of benefits for traditional working practices.

The emotion detection and recognition market is predicted to grow to around $65 billion by 2023, which shows how highly investable it is, and why organizations need to take heed of the capabilities. While there are limitations and ethical considerations, it is expected that these will be addressed as the technology continues to develop. Although they shouldn't be ignored, firms must focus on the potentially lucrative drivers of emotional intelligence.

Call
Toll free: +1 877 239 3153

Email
info@crypteklabs.com
crypteklabs.com