Analysis and Comparison of Intelligent Personal Assistants

Oktay Bahceci[a,]

[a]*oktayb@kth.se*

Abstract

Intelligent Personal Assistants (IPA) are implemented and used in Operating Systems, Internet of Things (IOT), and a variety of other systems. Many implementations of IPAs exists today and companies such as Apple, Google and Microsoft all have their implementations as a major feature in their operating systems and devices. With the use of Natural Language Processing (NLP), Machine Learning (ML), Artificial Intelligence (AI), and prediction models from these fields in Computer Science (CS), as well as theory and techniques from Human-Computer Interaction (HCI), IPAs are becoming more intelligent and relevant. This paper aims to analyse and compare the current major implementations of IPAs in order to determine which implementation is the most developed at this moment in time and is contributing to the sustainable future of AI.

Keywords: Intelligent Personal Assistant; Statistical Learning; Natural Language Processing; Machine Learning; Artificial Intelligence; Human-Computer Interaction.

Contents

1. Introduction

Intelligent Personal Assistants (IPA) and their current software implementations exists in a range of applications, usually integrated in the Operating System (OS) from different developers and organizations, such as in personal computers, mobile computers and in *Internet of Things* (IOT). The IPAs are implemented in different programming languages and behave differently but all fall under the same branches of computer science and concern themselves with similar problems [1].

Natural Language Processing (NLP) and Human-Computer Interaction (HCI) have always been essential parts for these type of systems, and more recently as the systems develop further, Machine Learning (ML) and Artificial Intelligence (AI) have become essential for the functionality of IPAs [1].

Apple, Google and Microsoft are organizations that have their own implementations of these kind of systems and are continuously integrating their implementations across all of their devices and services [2, 3, 4].

1.1. Scope and Objectives

This paper will introduce theory, concepts and branches of Computer Science (CS) that are fundamental for building IPAs, followed by an introduction to some of the existing enterprise applications from three different companies, Apple, Google and Microsoft, in order to finally analyze and compare their behaviour. Thereafter, the current state and future state of these systems are discussed with predictions about their future behaviour and finally, the conclusion chapter deals with the current best implementation in combination with predictions about its future features.

2. Background

This chapter, namely the *background* chapter introduces the concepts, theories and models that are relevant to *Intelligent Personal Assistants*. The first part covers a few definitions of essential parts of AI, followed by the fundamentals of some statistical learning methods and finally an introduction and discussion of *Human-Computer Interaction* and its relevance to *Natural Language Procession* in the context of IPAs.

2.1. Artificial Intelligence

Artificial intelligence (AI) is a scientific field which strives to *build* and *understand* intelligent entities. Existing formal definitions of AI address different dimensions such as *behaviour*, *thought processing* and *reasoning*. The distinguishing between *human* and *rational* behaviour is often mentioned in the field. To create AI the two components intelligence and tools are required. The field of Computer Science have created such tools [1].

2.1.1. Agents

An agent is something, such as an entity that acts, from the Latin verb *agere*, which means *to do*. Computer programs always acts and does something, but an agent is excepted to do more. An agent is expected to be able act autonomously, analyze and adapt to its environment and change, persist over a longer time period and finally create, understand and pursue goals [1].

2.1.2. Rational agent

A rational agent is an agent such that for each action and task, the agent always acts in such a way to achieve the best outcome. When there are uncertainties, the rational agent is considered to achieve the *best* outcome [1].

2.1.3. Intelligent agent

An *Intelligent Agent* is a type of agent that is considered to be *rational*, and with a set of design principles, developers are able to create successful agents that make the system reasonable intelligent in order to complete certain, *intelligent* tasks. IPAs are software agents that act on the behalf of the user in order to complete tasks and provide information. The communication between the entity and the user is usually based on voice inputs or commands [1].

2.2. Statistical Learning

This chapter will include definitions of certain fields that are relevant for the creation of *Intelligent Personal Assistants.Artificial Intelligence*, *Machine Learning* and *Sentiment Analysis* are introduced together with its contexts and relevance to IPAs.

2.2.1. Machine Learning

Machine learning (ML) is a subfield of AI concerned with the implementation of algorithms that can learn autonomously [1]. Statistics and mathematical optimization provide methods and applications to the area of ML because of its strong connections with ML, where both areas aim at locating interesting patterns from data [5].

A major issue and drawback for the use of ML and classification models is the risk of over-fitting, which is when a learning algorithm overestimates the parameters in the training data. The opposite of this is *bootstrapping*, which is creating fabricated data with statistical models with the help of a small sample data set. [6].

2.2.2. Natural Language Processing

Natural language processing (NLP) is a field in artificial intelligence and linguistics concerned with interaction between computers and human natural language. As a part of *Human-Computer Interaction*, NLP is concerned with enabling computers to derive and interpret human natural language. Recent work in NLP are algorithms based on ML and more specifically statistical machine learning [1, 7].

State of the art applications of NLP consist of text classification, information extraction, sentiment analysis, machine translation and is applied to many different scientific areas [7]. Discussed in depth in the next section, *Sentiment Analysis* approaches have been applied to IPAs.

2.2.3. Sentiment Analysis

Sentiment Analysis (SA), or *Opinion Mining* (OM) is the use of NLP, text analysis and computational linguistics to identify and extract information (or *features*) from texts [8].

In its current stage, automated SA is not able to be as accurate as human analysis. The automated sentiment analysis methods do not account for subtleties of context, environment, irony, human body language or tone. In human analysis, the inter-rater reliability plays a significant part, which is the

degree of agreement among raters. According to recent studies, the human agreement rate in sentiment analysis is around 79-80%. [9, 10, 11]. No IPA in enterprise form is considering any of the mentioned aspects and will therefore not be discussed further.

2.2.4. Supervised Machine Learning

Supervised machine learning aims to predict output data sets $(y_1, y_2, ..., y_n)$ from given sets of input data $(x_1, x_2, ..., x_n)$ for n observations. A general machine learning function is created for predicting output from the input that has not been a part of a training set. The predictions are formed by a training set of tuple data, such as, $((y_1, x_1), (y_2, x_2), .., (y_n, x_n))$, from a known set of input and output [1].

2.2.5. Unsupervised Machine Learning

Unsupervised machine learning is the process of classifying data without access to labelled training data. Using n observations of data $(x_1, x_2..., x_n)$ the primary goal of the unsupervised machine learning method is to gather data with similar attributes and relationships into different groups. As labelled data is not provided, unsupervised methods usually require larger amounts of training data to perform equally as good as supervised machine learning methods [1].

2.2.6. Artificial Neural Networks

In machine learning and cognitive science, *Artificial Neural Networks* (ANNs) are a family of models inspired by biological neural networks, more specifically the human brain. Artificial neural networks are commonly constructed in layers where each layer plays a specific role in the network and contains a number of artificial neurons. Typically these layers are the input layer, the output layer and numerous hidden layers in between. The actual computation, processing and weighting of the neurons is done through the hidden layers and is crucial for the performance of the network [12].

It can be speculated that most entities and IPAs take use of various Artificial Neural Networks (ANN) or combinations of Artificial Neural Networks with other types of techniques, such as Bayesian regularized ANN. ANNs are preferred due to their ability to deal with nonlinear relationships, fuzzy and insufficient data, and the ability to learn from and adapt to changes in a short period of time [13].

3. Methods

This chapter describes used research methods and data collection approaches. Furthermore, the methods used for Sentiment and Data Analysis are described.

3.1. Literature Study

Through research in academic articles, digital articles, papers and books within the area of AI, ML and CS, the theoretical principles of the field of IPAs has been analyzed.

The literature used has been accessed via Google Scholar and the KTH Library database using relevant keywords such as intelligent personal assistants, intelligent agents, machine learning, sentiment analysis, artificial neural networks, human-computer interaction.

Finally, corporate and SDK information regarding the three companies of choice has been fetched through their official websites.

4. Analysis and Results

This section introduces the analysis and results of the methods and background introduced in the previous sections of this text. The scope of this text is restricted to three different companies, namely *Apple Inc.*, *Google Inc.* and *Microsoft Corporation.*

4.1. Apple Siri

Apple's Intelligent Personal Assistant is named *Siri*. Siri is Apple's IPA and knowledge navigator, and is currently an integral part of their operating systems, namely *iOS*, the operating system used for their smart phones and tablets, watchOS, which is the operating system for their smartwatches, namely *Apple Watch*, and for their television hardware, the Apple TV, *tvOS* operating system [3].

Currently, Siri is included on the iPhone 4S, iPhone 5, iPhone 5C, iPhone 5S, iPhone 6, iPhone 6 Plus, iPhone 6s, iPhone 7, iPhone 7 Plus, 5th generation iPod Touch, 6th generation iPod Touch, 3rd generation iPad, 4th generation iPad, iPad Air, iPad Air 2, all iPad Minis, iPad Pro, Apple Watch, and Apple TV [3].

Siri was introduced the year 2011 and as of early 2016, Siri supports 17 natural human languages. Currently, Siri can edit the state(s) of the

following native applications coming with the operating systems; reminders, weather, stocks, messaging, email, calendar, contacts, notes, music, clocks, web browser, Wolfram Alpha, and Apple Maps [14].

4.1.1. Background, Research Development

Little is known about Siri and its architecture, but there exist some knowledge about its background and internal structure.

Siri's primary technical areas focus on a Conversational Interface, Personal Context Awareness, and Service Delegation, which are all essential parts for an IPA. The speech recognition engine that Siri has is provided by Nuance Communications, which is a speech technology company [14].

Siri also has hard-coded responses, or *Easter eggs* for conversational and comic reasons, such as *What is the meaning of life* and *Who is your creator* [3].

4.1.2. Development Specific Details and Speculations

Siri is most likely still under re-factoring and feature and QA development is in continuous development. The area of Artificial Intelligence and Machine Learning has become more of a priority for *Apple Inc.* during the recent years, while Human-Computer Interaction has always been an integral part of Apple products and services [15].

22 April 2015 developers from Apple, and more specific, the Siri platform team announced at its Cupertino, California, headquarters that Siri is powered by Apache Mesos [16].

4.1.3. Apache Mesos

Apache Mesos is a distributed systems kernel, and according to the official website,

Mesos is built using the same principles as the Linux kernel, only at a different level of abstraction. The Mesos kernel runs on every machine and provides applications (e.g., Hadoop, Spark, Kafka, Elastic Search) with APIs for resource management and scheduling across entire datacenter and cloud environments [17].

Mesos provides the developers core functionality such as

- Scalability to 10,000s of nodes

- Fault-tolerant replicated master and slaves using ZooKeeper

- Support for Docker containers

- Native isolation between tasks with Linux Containers

- Multi-resource scheduling (memory, CPU, disk, and ports)

- Java, Python and C++ APIs for developing new parallel applications

- Web UI for viewing cluster state [17].

It can therefore be concluded that most of Siri's processing power, if not all of it, is heavily dependent on this backend, and is by implication heavily dependent on an Internet connection.

4.2. SDK

Siri uses the concept of **domains** in order to classify what the utterance belongs to and what the user intent is. With the release of iOS 10, Apple created the Siri SDK, namely *SiriKit* for developers, giving them the ability to implement Siri usage in their own apps. SiriKit support is divided into domains, each of which defines one or more tasks that can be performed. In order to support SiriKit, apps must support one of the following domains

- VoIP calling, Messaging, Payments, Photo, Workouts, Ride booking,

- CarPlay (automotive vendors only) and Restaurant reservations (requires additional support from Apple) [18].

4.3. Google Assistant

Google Assistant, previously named as *Google Now* and various of other names, is Google's personal intelligent assistant, and was released July 9, 2012. Google Assistant uses a natural language user interface to answer questions, make recommendations, and perform actions by delegating requests to web services, which makes the service heavily dependent on Internet access. Google recently created their own application for iOS with restricted features, mostly because of the OS and SDK restrictions iOS has for third party developers [2].

Little is known of Google Assistant's system, architecture and technologies used, what language its implemented in and what dependencies it has. Currently, Google Inc. has their IPA under closed development and there is no SDK or any tools existent for accessing its features. Therefore, Google's contribution to this area is irrelevant for this paper and will not be discussed further.

4.4. Microsoft Cortana

Cortana is Microsoft's intelligent personal assistant that is able to access and change reminders, recognize natural voice, and answer questions using information from Bing. Cortana is able to act as a personal assistant for stock applications with its operating system, but is dependent on Internet access for fetching information from Bing [4].

Currently, Cortana is supported for seven different languages; English, French, German, Italian, Spanish, Chinese, and Japanese.

Cortana is the only personal assistant under discussion that is cross-platform and has an open SDK, and is implemented on the following platforms; Windows Phone 8.1, Windows 10, Windows 10 Mobile, Microsoft Band, Microsoft Band 2, Android, Xbox One, iOS and Cyanogen OS [4].

Cortana was the only personal assistant out of the IPAs that are under discussion that presented an SDK for third party developers before the release of iOS 10 and *SiriKit*.

4.5. SDK

Microsoft has opened their SDK for Cortana for third party developers and gives much freedom for developers to integrate Cortana into their own applications, and even supports new actions. This section will focus on the features of the Cortana SDK.

4.5.1. Speech

Microsoft provides the following speech platforms and services for your apps.

Windows speech Windows speech is a set of UWP APIs that enable both speech recognition and speech synthesis across multiple languages on all Windows-10 based devices, including IoT hardware, phones, tablets, and PCs. Cortana on Windows uses these speech APIs.

Speech recognition Recognize real-time audio from the built-in microphone, from a source other than the microphone such as a Bluetooth headset, or from a file.

Speech synthesis Convert text into audio.

4.5.2. Cortana Actions

Microsoft lets third party users define *actions* which provides users with functionality from their apps, based on either explicit user requests or user

context. Actions are however restricted to Windows 10 Desktop and Mobile, and Android.

Developers can define their own actions from scratch, or select from the two predefined actions such as *ordering food* and *sending messages*. Examples of own actions include, but not limited to *Get nutrition info* or *Turn on the lights* [19]. This makes the Cortana SDK incredibly flexible and capable of handling a wide spectrum of intents and actions, which the other IPAs from Apple and Google does not.

However, developers need to *register* their actions, which can be done without any cost, and will most likely be reviewed by a developer working for the Cortana Team at Microsoft.

5. Discussion

This chapter will present an analysis of the results, discussion about the limitations, methodical constraints together with a conclusion and future work of this thesis. The implementational and computational limitations are discussed with focus on restrictions on time, data quantity and machine learning implementations. Finally, the conclusion of the found results are discussed with advice of future research in the areas.

5.1. Limitations

The current and popular IPAs from Apple, Microsoft and Google are almost all under secret and classified development, and therefore, the architecture and technoligies used by these companies are unknown. The exceptions are Apple, which recently released their *Software Development Kit* (SDK) with iOS 10 for their IPA Siri and Microsoft, which released its SDK for their IPA *Cortana*.

6. Conclusion

The purpose of this paper was to research the capabilities of some of the current IPAs that exist in the market today. It can be concluded that the most flexible assistant today is *Cortana*, which is available for all of the popular mobile operating systems, including *iOS*, *Android* and *Windows Phone 10 Mobile*. Measuring intelligence and technologies is not possible at this time, mainly due to the fact that the IPAs under discussion are under closed development. However, the most flexible agent is Microsofts *Cortana*,

which has the richest SDK and allows third party developers to customize and create their own actions for their agent. On the ethics side of things, Microsoft is still the winner. Microsoft has established an ethical, well-formed agent that is open for developers worldwide and does not blindly implement new features for their agent without confirming the new actions. This is widely talked about in *Superintelligence*, where a stable, common ground needs to be established for the AI in question before the public can start tweaking the source code for the intelligence. [20]. With the most complex agent, hopefully, Microsoft will be open to working with Artificial Intelligence groups such as **OpenAI** which aims to promote and develop friendly AI in such a way as to benefit, rather than harm, humanity as a whole [21].

6.1. Future research

It has been established that intelligent personal assistants are entities with many complex components that require skills in programming, statistics, machine learning, artificial intelligence and ethics.

The results of this thesis add to previous research and puts emphasise of statistics, machine learning, human-computer interaction and ethics.

Future research in this area could go deeper into the components that are necessary and essential for the development and creation of IPAs, the sustainability between the components, as well as building an entity with a rich SDK that allows for further extension in terms of the technical skills and competence required and what ethics to follow when doing so. When or if Google releases SDKs for its IPAs, it would be of great interest to compare these three IPAs and their respective SDKs in order to see what platform has come the furthest in the area, and by natural implication, the one that is contributing the most to the area.

References

[1] S. Russel, P. Norvig, Artificial Intelligence: A Modern Approach, 3rd Edition, Prentice Hall, 2009.

[2] Google search, `https://www.google.com/search/about/learn-more`, (Accessed on 03/11/2016) (11 2016).

[3] Apple, ios - siri, `http://www.apple.com/ios/siri/`, (Accessed on 03/11/2016).

[4] Microsoft, Cortana - meet your personal assistant - microsoft - usa, `https://www.microsoft.com/en-us/mobile/experiences/cortana/`, (Accessed on 03/11/2016).

[5] D. Hand, H. Manilla, P. Smyth, Principles of Data Mining, The MIT Press, 2001.

[6] J. Ticknor, A Bayesian regularized artifical neural network for stock market forecasting, Expert Systems with Applications 40 (14) (2013) 5501–5506.
URL `http://www.sciencedirect.com.focus.lib.kth.se/science/article/pii/S0957417413002509`

[7] Google, Natural language processing (2015).
URL `http://research.google.com/pubs/NaturalLanguageProcessing.html`

[8] S. Doan, M. Conway, T. M. Phuong, L. Ohno-Machado, Natural language processing in biomedicine: A unified system architecture overview, CoRR abs/1401.0569.
URL `http://arxiv.org/abs/1401.0569`

[9] A. Pak, P. Paroubek, Twitter as a corpus for sentiment analysis and opinion mining, in: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), 2010.
URL `http://www.lrec-conf.org/proceedings/lrec2010/summaries/385.html`

[10] J. Wiebe, T. Wilson, C. Cardie, Annotating expressions of opinions and emotions in language.
URL `http://people.cs.pitt.edu/ wiebe/pubs/papers/lre05.pdf`

[11] M. Ogneva, How Companies Can Use Sentiment Analysis to Improve Their Business.
URL `http://mashable.com/2010/04/19/sentiment-analysis/`

[12] S. Olatunji, M. Al-Ahmadi, M. Elshafei, Y. Fallatah, Saudi Arabia Stock Prices Forecasting Using Artifical Neural Networks, International Conference on Future Computer Sciences and Application (2011) 123–126.
URL `http://ieeexplore.ieee.org.focus.lib.kth.se/stamp/stamp.jsp?tp=arnumber=6041425`

[13] D. Das, M. Shorif Uddin, Data mining and Neural network Techniques in Stock Market Prediction: A Methodological Review, International Journal of Artificial Intelligence & Applications 4 (9) (2011) 117–127. URL `http://www.airccse.org/journal/ijaia/papers/4113ijaia09.pdf`

[14] Siri everything you need to know! — imore, `http://www.imore.com/siri`, (Accessed on 03/11/2016).

[15] Wwdc 2015 - videos - apple developer, `https://developer.apple.com/videos/wwdc2015/`, (Accessed on 03/11/2016).

[16] Apple details how it rebuilt siri on mesos - mesosphere, `https://mesosphere.com/blog/2015/04/23/apple-details-j-a-r-v-i-s-the-mesos-framework-that-runs-siri/`, (Accessed on 04/11/2016).

[17] Apache mesos, `http://mesos.apache.org/`, (Accessed on 04/11/2016).

[18] Apple, Introduction to sirikit, `https://developer.apple.com/library/prerelease/content/documentation/Intents/Conceptual/SiriIntegrationGuide/index.html`, (Accessed on 03/11/2016).

[19] Microsoft, Cortana actions, `https://msdn.microsoft.com/en-us/cortana/actiontypeofaction`, (Accessed on 06/11/2016).

[20] N. Bostrom, Superintelligence: Paths, dangers, strategies (2014).

[21] OpenAI, Blog, `https://openai.com/blog/`, (Accessed on 09/11/2016).