Chatbot implementation as customer service support for an entertainment company

Adrià Riera Llambí
Master's in Business Intelligence and Big Data
Area of expertise: Business Analytics

Rafael Luque Ocaña
Joan Melià Seguí

07.2019

EXTRACT

EXTRACT

FINAL WORK'S SHEET

Title:	*Chatbot implementation as customer service support for an entertainment company*
Author's name:	*Adrià Riera Llambí*
Consultant's name:	*Rafael Luque Ocaña*
PRE's name:	*Joan Melià Seguí*
Delivery date (MM/YYYY):	*06/2019*
Title or program:	*Master's in Business Intelligence and Big Data*
Area of expertise:	*Business Analytics*
Language:	*English*
Keywords:	*Ruled-based chatbot; learning model; IBM framework; API*

Abstract (250 words or less): *With the purpose, context and applicability, methodology, results and conclusions*

Artificial Intelligence has proven to be a technical field with a big potential impact on the business strategy of a company. Concretely, chatbots could become the next technology to rule the customer services. They offer a gate to capture structured customer data for a posterior postprocessing and analysis, reducing the costs of operation compared to the traditional customer service tools. As well, among other capabilities, chatbots could have a better capacity to capture a loyal customer.

The pursuit of the present thesis was to create a chatbot that interacted with a user in a theoretical entertainment company from the film industry. Such company, having seized the market tendencies, would deploy a chatbot with basic functionalities as trial before a complete integration into the customer service department. Hence, it was initially designed to support the user during the search of product information and basic recommendations.

For such a purpose, a theoretical working environment and a product database for the company were defined. The rule-based bot was developed within IBM Watson framework. Its architecture and language were built considering the criticality of being a direct customer interface.

The final step was to characterize a learning and training model for the bot, as the machine learning engine available within IBM's framework did not cover directly all the bot's needs.

The realization of the thesis concluded with the online deployment of the bot and the identification of further potential developments that would reinforce the

impact on both the user's and company's side.

Resum del Treball (màxim 250 paraules): *Amb la finalitat, context d'aplicació, metodologia, resultats i conclusions del treball*

La intel·ligència artificial ha demostrat ser un camp amb un gran impacte en l'estratègia de negoci d'una empresa. Concretament els chatbots poden ser la propera tecnologia que domini l'atenció al client. Aquests ofereixen una porta per capturar dades d'usuari d'una manera estructurada (per un posterior post processat i anàlisi), reduint els costos d'operació si es comparen amb els de les eines d'atenció al client tradicionals. Entre altres capacitats, els chatbots podrien tenir una millor capacitat per capturar un client lleial.

La present tesi persegueix la creació d' un chatbot que interaccioni amb un usuari d'una empresa teòrica del camp de l'entreteniment cinematogràfic. Aquesta, després d'haver captat les tendències del mercat, desplegaria un chatbot amb funcionalitats limitades com a prova abans de la seva completa integració als processos de l'empresa. Per tant, aquest es dissenya inicialment per donar suport a l'usuari durant les tasques de cerca d'informació i recomanació de productes.

Així, es defineix un entorn de treball teòric i una base de dades de productes de l'empresa en qüestió. El bot és *rule-based* i es desenvolupa en IBM Watson. La seva arquitectura i llenguatge tenen en compte la criticalitat de ser una interfície directa del client.

L'últim pas és la caracterització d'un model d'aprenentatge i entrenament per al bot, ja que el motor d'aprenentatge disponible al marc d'IBM no cobreix totes les necessitats d'aquest.

La tesi conclou amb el desplegament en línia del bot i la identificació de possibles millores que podrien reforçar l'impacte en l'usuari i la companyia.

EXTRACT

1 Introduction

1.1 Background and rationale

A chatbot (also conversational bot or chatterbot) is an intelligent computer application capable of simulating the conversational skills from a human.

The idea of a computer program that would be not only able to communicate with a human counterpart but to do so in an (at least apparently) intelligent way was already conceived since the beginning of the history of artificial intelligence (AI). In the year 1966 the ELIZA program was developed. ELIZA was capable of mimicking a human conversation: it related the user entries with a series of keywords that accordingly triggered the corresponding answers [1]. Working mainly on the same premises, lots of different and more advanced versions of chatbot have emerged ever since. Thus, the program ELIZA was followed by others like PARRY in 1972, an evolution of ELIZA that simulated to be an individual afflicted with schizophrenia. Its initial version from 1981 (that has been evolving until the present days) was the first chatbot that could speak about general topics; newer versions also learned from the interactions with the different users, building up a dialogue background for the different interlocutors that made the conversations appear as more human alike. In 1995 A.L.I.C.E. was able to fluently hold a normal conversation based on heuristic techniques (evaluating the most suitable answer from a group of different options). Another example is IBM's Watson which in the year 2010 could win human opponents on a regular basis in the popular TV show Jeopardy! A schematic graphic evolution of the milestones in the development of the technology can be found under [2].

On the other hand, the democratization of computer applications and the appearance of a general computer literacy at the beginning of the 90's have been a catalyst that facilitated the embodiment of the first theoretical works about automatic chats into real life applications. Already in 1996 Microsoft included the Clippy help system as part of its Office Word consumer computing package. While Clippy was strictly speaking not a chat tool (mainly due to its limited discussion topics), it did use some of the new techniques that emerged in parallel to chatbots: based on the behavior that a user had during the work session and the information that was collected in the databases of Microsoft, Clippy detected if help in some specific task was needed and interacted accordingly. Its success was limited (from 2002 onwards the assistant was deactivated by default and from 2007 it stopped being included in the Office software [3]) but also turned out to be one of the first examples of tangible applications of such technologies for a standard consumer.

This line of development has not stopped growing. Currently, chatbot applications integrate new paradigms of artificial intelligence such as machine learning which is the ability of a system to learn from its experience, that is, to modify and optimize its operation with its use (adapted definition of the Oxford dictionary[1]). Here a small remark should be made: nowadays, the concept of *system* is preferred in the different literature rather than other terms (such as a computer or application) because it captures better the range of technologies that today integrate this type of solution, where the couple computer – program is only a part of the possible configurations. Being a system does not only improve the experience and interaction with the user but also makes them almost autonomous. Namely there are solutions where chatbots are integrated to instant messaging tools such as WhatsApp, Telegram or Facebook Messenger. The customer can take these channels to contact a company and get

[1] Available for free online.

automatically answered questions about the features of a specific product, resolve doubts, learn a language[2] or manage reservations. Similar features are available for mobile assistants. In a similar way, chatbots are also integrated as add-ons for navigation on web pages, where visitor assistance is provided as a guide through the page or to solve frequently asked questions. Another field of application is the voice-targeted attendees, such as Siri or Alexa, who are programmed to facilitate the daily work: tasks as organizing the agenda, ordering products (groceries, clothing…) or centrally controlling the devices of the house connected to them can be managed directly chatting with the appliances.

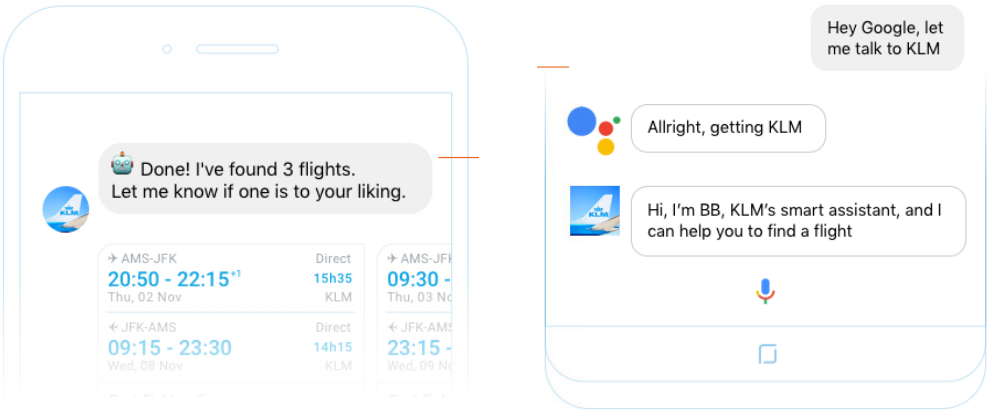

Figure 1: Captions of KLM's assistant Blue Bot; it can be called via Facebook Messenger (left) or Google Assistant (right). Source: KLM

On a business strategy level, the commitment to tools and resources based on AI is an advantage not only for the value that is extracted from the sale of the product itself but also because of its impact on the company. Specifically, in the case of chatbots, the expected impact can be summarized in the following points (mix from different sources and own contribution, see [4], [5], [6], [7] and [8]):

- Low maintenance, implementation and operation costs
- Customer service can be offered 24/7 with quicker response times
- Customer service can be available in different communication channels
- Technology can coexist with a more individualized customer service whenever automatization does not offer the right answer
- Implementation of chatbots also opens up the possibility to put the focus of human capital efforts on tasks of greater added value
- Possibility of complementing the offer of products with recommendations of similar products or services that are more accurate and adjusted to a specific customer profile
- Improvement in the quality of customer service (as consequence of the previous points)
- Low learning curve for new users
- Ability to manage parallel conversation threads (with customers, intermediaries, internal departments ...)
- Ability to present non-structured information (for example scattered through different databases or files) to the user in a simple and direct way

[2] For further info see, for example, Duolingo's offer.

- Easier collecting of organized customer data that could lead to a richer database for a later in-depth analysis
- Communication interfaces are reduced, as the communication channels are much simpler, immediate and usable on different platforms (mobile, internet, application, web page complement ...)
- Simple introduction in the company's processes

As introduced before, chatbot applications have an impact on both sides of the company processes. In this direction, the technology consultant *Gartner* classifies the current applications of chatbots in three basic categories [9]:

- Virtual Customer Assistants (VCAs), responsible for automating the written and voiced communication of the call centers
- Virtual Enterprise Assistants (VEAs) business assistants that are intended to simplify internal communication and employee access to information
- Virtual Personal Assistants (VPAs) responsible for managing third party services such as emails, social networks or making reservations for business trips

Gartner also estimated that in 2017 only 2% of companies included chatbot applications in customer services, while it is expected that by 2020 the number will have increased up to 25% [10]. This trend will also coincide with the abandonment of the mobile applications that many companies will undertake during the same period. These are allegedly expensive, inefficient and fail to capture a loyal customer [11]. Due to the features listed above, chatbots can occupy part of the technological and business gap that will appear in the future.

It is within this framework that the present work is carried out. Its goal is to create a chatbot that interacts with both, an internal (employee) and external (customer) user. Taking a generic company in the field of the film entertainment industry, the chatbot proposed will:

1) Allow the user (employee or customer) to consult the product database in a simple way (query function)
2) Recommend personalized products adapted to a specific customer taste (recommendation function)

The recommendation function is introduced to support the business strategy of the theoretically defined company. Companies count with recommendation systems to both gain new customers and to retain them. That translates into a revenue growth: some sources speak of an impact between 20% to 30% [12]; according to McKinsey, 35% of consumers' purchases on Amazon and 75% on Netflix come from their recommendation engines [13], the latter engine being allegedly worth 1 billion dollars per year [14]. Further expected impacts on the business are an improved customer knowledge that would derive to a more efficient inventory or logistic processes or a better customer-oriented experience.

The advantages of recommendation systems complement and enhance those of a chatbot. For that reason, a recommendation system is integrated as part of the current chatbot functionalities, boosting the positive customer experience and the added value of the technology. In the present work, it is intended to illustrate in a practical way the applicability of this type of platforms. Concreter objectives, approach, and schedule are detailed below.

EXTRACT

1.2 Objectives

This master's final thesis pursues the following objectives:

1) To define a theoretical working environment for the chatbot
2) To define a working database
3) To define a framework to develop the chatbot application
4) To create a chatbot-type solution that works with this database:
    a. To automatically answer questions about it (query function)
    b. Based on the user inputs, to recommend products that are appropriate and adjusted to his or her tastes (recommendation function)
5) To define the parameters and learning model for the bot

1.3 Approach and methodology

The high demand for products and computer applications and their constant change in requirements (in paradigms, technologies, calculation capacities, platforms ...) makes it a common practice for many companies to contact external suppliers of computer solutions, as the option to build internal competences is in many cases impossible: the acquisition of the needed knowledge and its material maintenance is too expensive and, in most cases, far away from the center of the competences of the company in question.

The case of chatbots is no exception: a quick search on the internet results in different models of support for companies that are interested in their implementation. Whether an integral cloud solution or do it yourself solutions are wished, different chatbot models for all levels of complexity are, to a greater or lesser degree, available online.

Within the actual project, due to the limited time allotted and the technical competences of the field of studies, a chatbot will not be completely developed from scratch. Nor is the pursuit to perform a comparative study of pros and cons of the most popular existing solutions. This project aims to take advantage of some of those existing solutions, more concretely frameworks, as base of the chatbot creation (see the chapter 3 Creation of the chatbot and analysis of the results, for more information). Hence, the structure of the bot constituted by the architecture, integration of external knowledge, the understanding of human counterpart, corresponding actions or supporting algorithms will be authored explicitly for the present thesis.

Therefore, first the characteristics of the working environment and the needs for a framework will be defined. Those will be the spine of the bot. After that, the most appropriate strategy to adapt the technical and business sides will be described. This way, the difficulties arise from the definition and resolution of the problem and not vice versa. This aspect is essential, for example, during the learning phase of the chatbot. How to tackle the different steps will be discussed in more detail in subsequent chapters.

1.4 Schedule

The work planning includes both the resolution of the previously defined objectives and the different deliveries and milestones fixed by the UOC.

EXTRACT

A time schedule can be found under chapter 7.1 Flowchart. The milestones are summarized as follows:

PEC01: Initial proposal	25.03 – 05.04
PEC02: Evolution and memory of the work	06.04 – 04.06
2.1 Extension of bibliography	06.04 – 12.04
2.2 Compilation of current solutions	06.04 – 28.04
2.3 Definition of the working environment	13.04 – 21.04
2.4 Database preparation	13.04 – 21.04
2.5 Creation and training of query chatbot	21.04 – 15.05
2.6 Creation of recommendation chatbot	15.05 – 04.06
2.7 1st draft report	15.05 – 04.06
PEC03: Final delivery	05.06 – 27.06
3.1 2nd draft report	05.06 – 22.06
3.2 Training of recommendation chatbot	05.06 – 22.06
3.3 PowerPoint presentation	21.06 – 27.06
PEC04: Evaluation committee	03.07 - 10.07
4.1 Evaluation of the thesis	03.07 – 10.07

1.5 Summary of obtained results

With the completion of this work, the following products are expected to be obtained:

1) Definition of a working environment
2) Database that simulates a real working environment
3) Definition and characterization of a learning and training model for the chatbot
4) Definition of the architecture of the chatbot based on the available data, the working environment and the query and recommendation functionalities
5) Functional chatbot
6) Code, files and / or other applications that can support the abovementioned points

1.6 Short description of the further chapters

The work is structured in different blocks. After this introduction, a study of the state of the art regarding chatbot solutions, reinforcement learning methods and recommendation models is provided. In the next step the project framework will be defined and analyzed in terms of:

- Working environment
- Architecture and platform(s) used for the chatbot creation
- Chatbot working principles
- Chatbot actions, such as:
  a. Query
  b. Recommendation
  c. General interactions with the user

EXTRACT

- Learning model

The results obtained from the process will give way to the conclusions. The last chapters introduce the lexicon of terms and abbreviations appeared during the text, the bibliography and the annexes.

EXTRACT

4 Conclusions

At this point of the document, once the chatbot is active and deployed, is time for a recapitulation to compare the initial goals of the thesis stated in 1.2 Objectives against the actual results obtained.

Along the chapter 3.1 Working environment, the context of the entertainment company, an application field for the MovieBot and a database were defined and with that the two first objectives. The internet will act as information database of the bot, as no product catalogue was available, nor one was going to be created (the later not being in the frame of the thesis); two products will be disposed for consultation: actors, actresses and directors or films. Information on the first products will be taken from the Wikipedia; for the films, two free of access APIs will be used: OMDB and TMDB. The first considerations can be extracted:

1) The bigger risk of the implementation of the chatbot is on customer side, so keeping user experience will vertebrate the complete bot architecture.
2) The theoretical environment is a general case that introduces restrictions to the bot architecture. Any real environment will be a reduced case of the defined one.

Being the implications of the second point explained along the next subchapters.

The bot is developed within the IBM Watson framework. IBM Watson framework provides a set of tools, learning engine, and easy training of a rule-based dialogue, what avoids creating the bot from scratch. The framework showed several advantages when compared to its concurrence: the free version allows an easy integration of external functionalities (very relevant for the communication with the external APIs) and the building up of complex dialogue flows, while being its deployment uncomplicated and compatible with different communication channels. This points together with the architectural elements that build the dialogue are extensively described in 3.2 Framework.

In 3.3 Architecture description is treated the fourth objective, the creation of the chatbot itself. Following the considerations previously extracted, the subobjective of offering a quality service to the user is also embraced during the process. The conversation of the bot consists of an overlapping of condition-action couples, grouped into nodes. A simplification of the resulting architecture is given in the Figure 6. A practical interaction with the bot can be run via its online deployment or directly in the framework.

Among others, the architecture of the conversation successfully confronted several issues:

3) The bot needed to show a certain personality in its answers, aligned with the image of the company it represents. To strengthen the image with a certain social intelligence, the bot was enhanced with some small talk capabilities and with a transfer function to another communication source.
4) The bot had to simulate a certain level of intelligence in terms of capacity to correctly understand the user, to control the conversation flow and to know its own limits. That translated into a language structure and architecture that:
    a. Uses concrete and clear phrasing and vocabulary, avoiding long sentences and formulations that may sound strict, cold or rude.

b. Has always an active control of the discussion flow in order to bring it to the main competences of the bot, to avoid getting in repetition loops or to manage unclear inputs from the user.

c. Tends to nudge the user to the core competences of the bot.

d. Is able to correctly classify the intent of a user input what includes having a certain flexibility to react to the change of mind of the user.

e. Keeps a simple and understandable dialogue evolution as the bot is planned to interact with different profiles and skills.

f. Inserts information extracted by the bot from the conversation in the answers to show understanding and reinforce the simulation of intelligence.

g. Analyzes the validity of certain user inputs.

5) Per se, the bot is incapable to connect with the external APIs. In order to fulfill the recommendation and query functions, the bot is integrated with two cloud functions that are the active responsible to call the information from OMDB and TMDB and pass it back to the bot.

The Figure 13 resumes the results obtained after the successful call of the query functions; the Figure 15 the ones from the recommendation function. The recommendation method introduced is content-based, that primarily filters out the data (in this case a list of films) based on two genres chosen by the customer. The recommendation system was driven by the technical and temporal limitations. Being the thesis developed within a data analytics framework, the opportunity to schematize in a very global way a second recommendation method based on reinforcement learning was not dribbled. The functions were described in detail in 3.3.4 Query nodes: Query Mode and Define Product Type and 3.3.5 Recommendation nodes.

All the previous conveys to the fifth and final objective, the characterization of the learning model for the bot. The objective is covered in 3.4 Learning model. The understanding of the bot is simulated via its capacity to correctly allocate a user input to the corresponding defined intent. Each of the intents is associated to a set of triggering sentences, words or parameters; after that, the bot is automatically trained to understand deviations from the triggering data via machine learning. In case the automatic training is not working properly, the framework also includes the possibility to manually train the bot; an example is given in the Figure 16. For the MovieBot case, the complexity arises when the bot has to locate the product name inside the user's answer. Product names are non-standardized inputs (they do not have a unified format, length or structure) nor are pared to any IDs, so the engine is not capable of correctly capturing them. As a side note, in this direction Google's DialogFlow proved to have a better performance in this aspect; as these concrete points are outside the scope, the comparison between both frameworks is attached in the chapter 7.6 DialogFlow vs IBM Watson from the annex.

The limitation is solved via architecture: idea is to accept any input from the user as a potential product name, validate the intent of the input via filtering (see Figure 17) and then, for movies, check against the API if the title exists. As for actors, actresses and directors there is no database, the final check is not done for that kind of products. Being most of real-life applications a particular case of the MovieBot, and as a solution is implemented and working for the movies, the issue is considered to have a low impact in the final performance of the bot, as a real case would allow an immediate correction.

3.4.2 General considerations contains the set of key elements for the training of the MovieBot, being the most important of them:

6) The bigger the set of training examples, the more precise are the model and the capacity to correctly classify user inputs; that includes a rich formatting of training sentences in terms of length, structure or vocabulary, also including typos.
7) IBM Watson's engine is not capable of extracting new entities from a training set, nor to segregate them from a sentence.
8) Using synonyms for entities improves the bot's performance when capturing important user information; this does not counteract the previous drawback.
9) The use of keywords in the training set (concepts that can be indistinctly associated to an only intent) reinforces the understanding capabilities of the bot.

In the introduction to the work, it was justified that there is a market potential in the deployment of chatbots and that those chatbots can be the spark to a long-term business strategy. Chatbots could be initially used as customer interface and as entry to future business developments or other areas of the company. That approach placed the biggest risk on the customer, so the bot was built to have certain social skills and to use easy language and conversation flow, to adapt to different user profiles. The recommendation function was included to increase the impact of its deployment on the company's revenue; the same impact in the revenue could potentially subsidize part of the investment needed for the test phase of the deployment.

The final pursuit of the thesis was to show in a practical and cost-effective way how a chatbot could be deployed and support a theoretical company. It can be concluded from the above points that the pursuit is satisfactorily achieved.

For that, some modifications in the initial scope of work and project planning were introduced to adapt them to the evolution of the project and as the first results were obtained. The biggest redirection was in the project objectives. The first idea was to center the work around reinforcement learning methods. After some research in the area, the approach turned out to be too wide and technically far from the core competences of the field of studies (business analytics). That justified a change of focus towards more practical applications of derived methods as machine learning supported chatbot applications.

The difficulties arose from the limitations in technical knowledge were also palpable after the first bot models were created and the need to enhance their basic functionalities via external APIs was faced. Some learning in JSON and Java coding was necessary to cope with that phase, as well as in the internal working methods of the different frameworks. That derived into the change from DialogFlow to IBM Watson treated in the annex what also impacted the scheduled plan.

Even after the new project objectives, it was considered that the influence of reinforcement learning in the bot development was still slightly present along the bot training phase. Reinforcement learning methods showed also a big potential in a future improvement of the bot. For those reasons, a short resume in that aspect has been kept in the chapter 2.2 Reinforcement learning. That fact is a good link to step into the anteroom of the end of the conclusions, the open points.

On a technical level, the following points were not completely developed:

i) Rule-based conversation always has the counterpoint that reduces an infinite problem to a set of limited condition-reaction pairs; that could be improved using reinforcement learning techniques that would pick up between different answer options.

ii) For the same reasons as stated above, in some cases, bot's answers may appear to be a little bit cold and out of the blue. That could be improved by working on a better formulation of bot's answers and including more nodes outside the core competences.

iii) A better recommendation model can be implemented following the explanation given in 3.3.5.1 By genre and popularity and 3.3.5.2 Other recommendation methods.

iv) A user-oriented recommendation model or even bot interaction could be also implemented. The bot would adapt its answers, language or even the conversation style to the user profile.

v) If the product name corresponds to a movie saga, the API provides directly the first of the list. To avoid these cases, a disambiguation capability (similar to "Did you mean…?") would be necessary. That could also redirect misunderstood user intents.

The potential of the bot could be enhanced with further functionalities. The ones considered to have more relevance are:

vi) The bot could be used as a gate to gather consumer information for further offline customer profiling. That would be useful to study consumer tendencies, common doubts or to feed a better recommendation system.

vii) Actively integrate a FAQ mode to tackle most basic questions regarding the bot functionalities, products and also regarding the company.

viii) Being an international company, the bot should incorporate other languages than English.

ix) Bot should be able to react to voice interaction as well.

x) Bot should be deployed in all the communication channels from the company it represents.

xi) Bot should be integrated with human-driven user attendance.

Even when the real impact of the bot on the quality of customer services or its cost would need to be specifically validated (interviewing customers, defining test and control groups…), the current design of the MovieBot has a degree of maturity and of consideration of the customer needs that makes it ready for a first trial deployment without the need to include any of the open points listed.