A Conversational Chatbot Architecture for eHealth Systems

Niccoló Terreri, Master of Science in Computer Science candidate

Contents

Chapter 1

Introduction

> "Pointing will still be the way to express nouns as we command our machines; speech is surely the right way to express the verbs."
> **Frederick P. Brooks Jr., 1995**

The Problem

According to the 2014 National Cancer Patient Experience Survey National Report, over the past couple of years only slightly over 20% of cancer patients across the UK reported having been offered an assassment and care plan specific to their personal circumstances (Quality Health, 2014, p. 114). In an effort to increase the number of cancer patients who received such assessments, Macmillan Cancer Support piloted the Holistic Needs Assessment (HNA) questionnaire and health plan in 2008 (Macmillan, Holistic Needs

Assessment). This is essentially a self-assessment questionnaire where the patient identifies what their concerns are from a range of personal, physical, emotional and practical issues they may be facing in their lives in relation to their condition. The completion of the questionnaire is followed by the creation of a care plan through a consultation with a clinician, with further advice and referrals as needed. Macmillan began trialing an electronic version of the questionnaire in 2010, progressively extending provision of the eHNA to more and more sites (clinics etc) (Mac Voice, 2014).

Intelligent conversation systems have enjoyed an increasing amount of media attention over the last year[1]. With applications of artificial intelligence to using natural language inputs for different purposes, including general purpose mobile device interfaces[2]. Furthermore, several technology companies have started offering "Artificial Intelligence as a Service" products. Among these are BloomsburyAI (founded at UCL) and bespoken companies such as Google and Microsoft[3]. This appears indicative of the fact that chatbot and natural language processing technologies have reached a level of maturity comparable to that achieved years ago by haptic technology, that we find almost ubiquitously in human-computer interfaces and everyday use of computing devices today.

This project is a about the use of an intelligent conversational system to gather further information about the patient's concerns through an electronic self-assessment tool, ahead of the creation of a patient care plan. This

[1] Numerous articles, among which: (The Economist, 2016), (Berger, 2016), (Knowledge@Wharton, 2016), (Finextra Research, 2016)

[2] (Viv, 2016), (Dillet, 2016)

[3] (Pandorabots, 2016), (Bloomsbury AI, 2016), (Microsoft Cognitive Services, 2016)

is primarily an attempt at introducing the conversational User Interface in electronic health applications generally, and in particular explore the applicability of computer advisors to Macmillan's eHNA in a growing effort to improve the quality of support cancer patients receive across the UK.

The scope of the present report is limited to the architecture and implementation of the chatbot system, as opposed to a complete, user-facing product: the complete application is a joint effort of four, with distinct concerns being assigned to different members of the team. The author of the present document being tasked with design and implementation of the core system backend. The other members of the chatbot team include: Andre Allorerung (MSc SSE) as the technical team lead who also takes care of the integration of the system with the resources available to PEACH and the data storage system that will persist durably information gathered through the chatbot system. Rim Ahsaini (MSc CS) working on a specialized search engine for resources that may interest and help support cancer patients based on their concerns (to be available both through conversation with the chatbot and independently), Deborah Wacks (MSc CS) as lead UX designer working on the implementation of a webserver through which allow the user to interact with the chatbot and search engine.

This project is part of PEACH: Platform for Enhanced Analytics and Computational Healthcare (Project PEACH, 2016). PEACH is a data science project that originated at University College London (UCL) in 2016 that sees Master level candidates working together on the data platform and on related projects. With more than twenty students across multiple Master courses, it is one of the largest student projects undertaken in recent years at UCL, and it is part of a long-term strategy to bring the UCL Computer Science department and the UCL Hospital closer together.

Project goals and personal aims

The main project goal is the delivery of a basic but easy to extend and modify chatbot software system, specifically targeted at assisting with the identification and gathering of information around cancer patient issues, modelled after the Concerns Checklist (CC) electronic questionnaire form (NCSI, 2012). Finally, one of the major challenges with eHealth problems is represented by having to hand confidential patient data (as will be discussed in Chapter 2 of this report). Summarily:

- Design and implement a chatbot architecture tailored to the issues surrounding software systems in healthcare (in particular around treatment of sensitive patient data)
- To integrate with a specialized search engine (developed by another member of the team)
- To explore other applications of NLP that could be useful to extract information from natural language data.
- To implement a chatbot brain using open source technology.
- To develop the system with Macmillan eHNA as the main reference.

Personal goals of the author include:

- Learning Python in an effort to gain exposure to a new programming language
- Leverage the author's background in computational linguistics, and explore the field of natural language processing
- Learn about applications of machine learning to natural language processing
- Improve software engineering skills by applying best Agile practices

The project approach methodology

An Agile approach was adopted for the project, in line with the author's stated interests. This meant maximizing time spent outside of meetings, save for where communication between team members and others was required. The project was paced in weekly iterations where aspects of the system to implement would be selected from a backlog to be delivered for the next week. Great emphasis was also put on testing as part of deveopment, in particular the discipline of Test Driven Development.

A top-down system design and implementation was also adopted, with the next largest system abstraction being priorized first in development in order to always have a working system being progressively refined. These methodology guidelines where established in accordance with the reccomendations of Brooks (1995, pp143-144, 200-201, 267-271) and Martin (2009, pp121-133; 2003, chapter 2, 4, 5).

Report overview

This report is structured as follows:

- Chapter 2 provides more extensive background into the eHNA questionnaire as well as NLP and chatbot open source projects that were explored.
- Chapter 3 describes the requirements as gathered through the contacts in healthcare and the Macmillan charity available to PEACH.
- Chapter 4 details the system architecture, design and the current implementation, highlighting its current limits and its extensibility.
- Chapter 5 discusses how system testing was done as part of development, the benefits of TDD to systems design and the evaluation of the machine learning component of the system.
- Chapter 6 concludes with an evaluation of the project results, a team retrospective and reccomendations for the direction of future work on the system.

Chapter 2

Background Research

> "Both the tractability and invisibility of the software product exposes its builders to perpetual changes in requirements."

The electronic Health Needs Assessment questionnaire

Macmillan Cancer Support

The eHNA system represents the background project against which the PEACH chatbot team efforts have kept constant reference to from the project inception throughout development.

Macmillan Cancer Support developed the eHNA for the purpose of extending the range of cancer patients in the UK covered by individual care plans, made with the individual's very personal and unique concerns they incurred into in relation to their condition. These concerns are gathered through variants of an electronic questionnaire offered by Macmillan to selected trial sites. Paper versions and variants of the questionnaire existed before the introduction of the eHNA in 2010, and have been in use since before then (Mac Voice, 2014) (NCAT, 2011).

The electronic questionnaire is designed to be carried out on site mostly through haptic devices (such as tablets), just ahead of meeting the clinician that will help draft a care plan for the patient. There is the option to complete the questionnaire remotely, although the adoption of this alternative is made difficult by the work habits of key personnel, who are used to providing a device to the patient in person and ask them to carry out the questionnaire while at the clinic.

The patient uses device touch interface to navigate through various pages selecting concern categories from a predefined list. There are several versions of questionnaires available, modelled after the various paper versions, depending on which one the clinic previously used.

Patients typically select three-four concerns (up to around six, mostly depending on the type of cancer they have). The questionnaire takes on average less than 10 minutes to complete. The information extracted is first stored in a Macmillan data store, external to the NHS N3 network (NHS, 2016). At this stage, Macmillan data storage synchs within a minute with data storages inside N3 and deletes all identifying patient information from the data is anonimized and data about the concern is retained by Macmillan to gather insight into the needs of cancer patients (consent is explicitly required from the patient in order to undertake the eHNA and information about the use of the data is transparently provided).

The front end of the system is implemented as web-app, built using HTML and JavaScript. Access to the assessment is restricted to scheduled appointments that clinics set up for individual patients, either via delivering the questionnaire on the clinic site, or, if the questionnaire is carried out remotely, via use of a one-time 6 digit PIN number, alongside the patient's name and date of birth.

The Concerns Checklist

Given the variety of different versions of the questionnaire, the team was advised to focus on the one that is most commonly used: the Concerns Checklist (NCSI, 2012).

In this version of the questionnaire, the patient selects their concerns from a range of more than 50 individual issues, each falling into one of 10 categories. Each category may itself be a subcategory of the following major topics:

- Physical concerns
- Practical concerns
- Family concerns
- Emotional concerns

- Spiritual concerns

Patient Data for Research in the UK

As mentioned in the project goals section in Chapter 1, handling confidential patient data poses particular challenges to eHealth related project. Just before the start of the project, when teams and roles had not yet been defined, the whole team underwent training about handling patient data and the relevant legislation in the UK (the specific certificates obtained by the author can be found in the Appendices [#MAKESURETHISHAPPENS]).

The following is a summary of key policies the author became familiar with before starting the project, with references to how in particular they affected design and implementation decisions.

Generally speaking, authorization must be provided before any information provided by the patient can be used in any way except the specific purpose of their healthcare. This severly limits the possibility of using third party services.

First, there is no guaranteed that the information can be transmitted securely to the external system. Secondly, this increases the risk of loss and inappropriate use of the information, both due to mishandling by the third party (whether intentional or accidental) and by increasing the risk that people unaware of the relevant legislation may come into contact with the data.

The Natural Duty of Confidence

Under UK common law, information that *can* reasonably be expected to be held in confidence under the circumstances (such as the information provided by patients to a clinician), *must* be held in confidence. This applies regardless of whether the information is specifically relating to the patient's physical health, and applies to any practical or other concern the patient may express.

Duties are sometimes contrasted with obligations in the sense that an obligation is a voluntary covenant a person enters, whereas a duty applies to the person regardless. This means that any personnel (including data scientists and software developers) who work with NHS patient data can be liable for misuse of the data even if they are not formally contracted.

The Data Protection Act 1998

The DPA (Data Protection Act, 1998) describes eight principles meant to ensure confidential data about (living) individuals is treated with fairness, and applies to any organization handling such data (e.g. financial institutions).

The nature of information covered by the act is "sensitive" in the sense that it may be used in ways that affect the subject to significant extents. Identifying information (such as name and date of birth) is normally regarded as such.

The second principle of the DPA specifies that the purposes for which personal data is being gathered have to be transparently described to the person.

This means that information provided by a patient for the purpose of their own health care can only be used for this purpose and no other (including mass aggregation of data to gather insight for any purpose from third parties involved).

The eight principle also requires that personal information is not sent outside the European Economic Area in most cases, which can also cause problems with the geographic location or accessibility across the world of data shared with third parties.

Conclusion

Patient consent should be gathered explicitly, having clearly explained all of the purposes for which the data may be used, before any information about them can be processed (with few exceptions, for example where the information becomes critical to national security and similar cases).

It may be possible to make use of third party services provided the data has been fully anonimized and cannot be linked back to the patient, and provided a special agreement (such as a Data Transfer Agreement) has been brokered to ensure both parties understand the legal and ethical implications of sharing even anonimized data; furthermore, it would be best to also gather consent explicitly from the patient even where the data has been anonimized. In such cases, the duty of confidence does not extend over to the third party. Note however, that it is sometimes difficult to ensure that data has been anonimized, even by removing all information considered personal under UK law: for example, if a person happens to have a rare disease, or information about the geographic location of the patient can be retrieved from the data being shared with the third party.

For this reason, the implementation of the current project does not share any of the data extracted from user input externally although the emphasis on clean architecture will allow for such a choice to be made in a future iteration of the chatbot project, where an agreement has been brokered, or authorization is otherwise provided to make use of third party services.

Natural Language Processing

As stated in Chapter 1, part of the author personal aims included to learn about NLP and leverage the author's background in computational linguistics. Furthermore, to investigate the application of machine learning to extract information from user input that would be relevant to the chatbot system as a whole.

To be specific, it is not so much tasks of information extraction or named entity recognition that were identified as most useful in this case, but rather the possibility to classify user input according to sentiment analysis and text classification. In order to inform the chatbot system reply to the user by providing additional information to the raw user input. Why text classification tasks? Because they may be used to add tags to user inputs to be matched within a chatbot brain that would otherwise be unable to generalize.

This would effectively represent a hybrid model where both a machine learning component and a “rule-based” (or rather input-pattern driven) approach are used as part of a more complex system.

In the abstract, this is an attempt at exploiting the power for generalization that is characteristic of approaches to artificial intelligence resembling the nature of experience as a way humans acquire knowledge, human capacity (and propensity) for inductive reasoning (Russell and Norvig, 1995, p.592; Biermann, 1986, pp.134-135; Hawthorne, 2014; Hume, 1777, section 4, part 2), while at the same time retaining the rigor and control provided by more rigid rule-based approaches to AI, which would ensure the conversation remains on the range of topics and allows the user to carry out the information gathering phase for the creation of their care plan, where a set of heuristics or rules of inference is used, these in turn resemble deductive reasoning, where certainty of the conclusion of the reasoning is guaranteed by the soundness of the deductive calculus used, and the mathematical certainty in the premises needed (Russell and Norvig, 1995, pp.163-165, Beall and Restall, 2014).

This approach is in particular to be contrasted with neural networks and other approaches to AI which behave as “black boxes”, and cannot by their nature proivde an intelligible explanation of their categorization process (or more generally decision process) as the result of their learning is stored in the form of a weighted graph (Russell and Norvig, 1995, p.567).

Text Classification

Text classification is the NLP task of assigning a category to an input from a predefined set of classes (Sebastiani, 2002, p.1). More formally:

For a document d and a set of categories $C = \{c_1, c_2, c_3, ..., c_n\}$, produce a predicted class $c \in C$

Particular to our case, the documents will be natural language conversational user input, and the set of categories will be the macro cateogries of issues that have been extracted from the concerns checklist (CC) version of the questionnaire (see above):

$C_{cc} = \{phyisical, practical, family, emotional, spiritual\}$

To be precise, we will be focusing on document-pivoted classification, where we wish to approximate an ideal mapping from the set of documents D to our $C_c c$:

$f : D \mapsto C_{cc}$

The task is normally turned into a supervised machine learning task, by training a model over a set of document-category pairs (Sebastiani, 2011, slide 7, 13):

$\{..., \langle d_i, c_j \rangle, ...\}$

Supervised learning is a form of machine learning where the machine is trained over a set of "hand" labelled examples. The "supervision" consists in already knowing the right answer for each training input, and wanting to use the system to automatically label future instances as desired (Russell and Norvig, 1995, p.528). This is in contrast with other forms of machine learning, such as reinforcement or unsupervised learning, where the answer is either unkown or "fuzzy" unlike with supervised learning. Take for example a robot navigating an industrial warehouse, as the surrounding circumstances change, the behaviour desired cannot simply be evaluated in binary terms (i.e. either "good" or "bad") because the problem of navigating a busy environment is by its nature not binary, there is in most cases a continuous spectrum of evaluation.

A model is trained over the training set and then tested against an unseen test set, also made up of hand-labelled samples. The model classifies each test sample and evaluation metrics can be drawn from comparing the model classification with the known ("gold") standard for the sample.

The internal representation of each document to the classifier is a sparse vector representing the features or characteristics of the document relevant to the classification task. Different features will be relevant to different document classificaiton tasks, for example certain words may occurr more frequently in positive movie reviews as opposed to negative movie reviews, but those particular words are unlikely to also be indicative of whether the person who wrote the document happened to be male or female.

These features can be, for example, the occurrence or non-occurrence in the document of certain terms (usually words).

Chapter 3

Requirements Gathering

Chapter 4

System Design and Implementation

Chapter 5

System Testing and Evaluation

Chapter 6

Conclusion