Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 198 (2022) 397–403

Procedia Computer Science

www.elsevier.com/locate/procedia

The 2nd International Workshop on Artificial Intelligence for Natural Language Processing (IA&NLP 2021)
November 1-4, 2021, Leuven, Belgium

Towards highly adaptive Edu-Chatbot

Tarek AIT BAHA[a,*], Mohamed EL HAJJI[a,b], Youssef ES-SAADY[a], Hammou FADILI[c]

[a] *IRF-SIC Laboratory , Ibn Zohr University, Agadir, Morocco*
[b] *Centre régional des Métiers de l'éducation et de la formation - Souss Massa, Morocco*
[c] *Conservatoire National des Arts et Métiers (CNAM), Paris, France*

Abstract

Conversational Agents are widely used in different domains to automate tasks and help to improve user experience. In recent decades, AI systems, thanks to deep learning methods and Natural Language Processing (NLP) approaches, can interact with users, understand their needs, map their preferences and recommend an appropriate action with no human intervention. However, chatbots in the education field have received limited attention. In this work, we use Xatkit, a chatbot development framework, for the definition of our Chatbot and propose an Encoder-Decoder framework for intent recognition. For the encoder, we encode utterances as context representations using bidirectional transformer (CamemBERT). For the decoder, we use an intent classification decoder to detect the student's intent. Our chatbot will be tested in the field of education to improve and simplify teaching for professors and learning for students as well as reducing faculty burnout and raising the speed of comprehension.

Keywords: Chatbot; Machine Learning; NLP; Transformers; Education

1. Introduction

Conversational AI bots are one of the many great promises of information technology. They were designed as a new interface to replace applications and supplement website visits by simplifying the interaction between services and end-users via chat. Usually, these bots are able to process Natural Language Processing (NLP) and provide answers to user questions. However, these responses do not always come in the form of text, but sometimes constitute actions, such as booking a flight, checking emails. Hence, certain industries, such as customer services and banking, are adopting this technology, allowing its users to use their systems so easily. There is another area where chatbots could have huge potential and that is education.

[*] Corresponding author. Tel.: +212 654509345.
E-mail address: t.aitbaha@uiz.ac.ma

10.1016/j.procs.2021.12.260

In a classroom, each student has different learning needs and interests. Therefore, everyone needs the help of a specialist tutor. Unfortunately, this type of service is not available even in the most developed schools around the world. Nowadays, it has become common that students use messaging services, which are standard features in platforms such as Google classrooms, and other class management systems, to communicate with each other and, occasionally, with their teachers. This kind of feature aims fundamentally to ask questions and obtain answers that allow students to build a better understanding and support the learning process outside the classroom. Using chatbots, this process could be replicated on a large scale, generating channels where students could discuss any topic with an ”expert”, ask questions, and reach conclusions that would improve their understanding of different topics. This process will ease tracking each student's improvement in real time.

In this paper, we used Xatkit [1], a development framework which provides a set of Domain Specific Languages to define chatbots. Xatkit proposes connections to different cloud-based Intent Recognition Providers such as Google's DialogFlow engine and IBM's Watson Assistant. In this work, we propose an in-house NLP Engine for Intent Recognition task. Our proposed model uses a learning approach based on the CamemBERT [2], and detects the intention behind the utterance through an Intent Classification decoder.

The rest of the paper is structured as follows: *Section 2* analyzes the existing models used to build a chatbot, and explores different architectures of Deep Learning and NLP approaches for intent classification task. *Section 3* presents the proposed system architecture and finally, *Section 4* concludes the paper.

2. Related Work

2.1. Chatbot types

In recent decades, there has been an increase in the development of conversational agent systems in different sectors such as E-commerce, banking, and education. In order to improve accuracy and make the user-bot conversations more realistic, chatbot development models have been developed from the traditional rule-based models to the retrieval-based models, then, to the generative models.

The traditional *Rule-based models* answer questions based on rules used in the training stage. These bots are created through a rule-based approach. However, These are not efficient in answering questions that do not match with the predefined rules.

The *Retrieval-based models* use predefined question/answer pairs. Then, they match user's queries against the predefined questions through simple algorithms like keyword matching or using more complex processing like information retrieval models. Next, they return the most suitable answers to the matched question as a response to the user's query. Since these models use a predefined pair of question/answer, they return responses with no grammatical errors. However, they have some shortcomings in that bot's responses are limited to the predefined set and are not sensitive to changes of queries.

To deal with these deficiencies, *Generative models* have been proposed. These models use Natural Language Processing (NLP) techniques and Deep learning techniques to model and train the Chatbot system. Most of these proposed models use the sequence-to-sequence approach that emerged in the Machine Translation, Speech Recognition, and Text Summarisation fields. Generally, Seq2Seq models consist of two recurrent neural networks (RNNs) called Encoder and Decoder. The encoder encodes the input sentences into a semantic representation by consuming the words from left to right, one by one. Then, the decoder decodes this fixed-length representation to generate the target sequence. The original Encoder-Decoder uses Vanilla RNN by default and has several drawbacks when long sequences are fed to the model. Since NLP depends on the order of words, it is useful to have a memory of the previous elements when processing new ones. Thus, it is beneficial to use the variation in RNN such as bidirectional RNN (BRNN), Long Short-Term Memory (LSTM), or Gated Recurrent Unit (GRU).

2.2. Chatbot frameworks

Nowadays, many development frameworks provide bot services where we can develop the bot and deploy it to any cloud. There are some cloud platforms that provide different services apart from the bot service such as built-in artificial intelligence, cognitive services, and so on. All these platforms are powered by machine learning algorithms. Table 1 presents some of them.

Table 1. Chatbot Frameworks.

Chatbot Frameworks	Services
IBM Watson[3]	Offers to build conversational interfaces into any application. It is built on a neural network, understands intents, interprets entities and dialogs. It supports English and Japanese languages, and provides different development tools such as Node SDK, Python SDK, and IOS SDK.
Microsoft Bot Framework[4]	Has its own Bot Builder SDK that includes .NET SDK and Node.js SDK. It supports translation to more than 30 languages and uses LUIS [5] for Natural Language Understanding and Cortona [6] for voice. To host the bot in an application or a website, the framework provides the Direct Line REST API.
Wit.AI[7]	Allows using entities, intents, actions, and contexts. It offers Natural Language Processing engine and supports about 50 different languages.
Rasa[8]	Provides a set of Machine Learning tools to build contextual chatbots and assistants. The framework consists of two components: Rasa NLU, which is responsible for Natural language understanding, and Rasa Core, which helps to create intelligent chatbots.
DialogFlow[9]	A Google AI chatbot framework that comes with Machine Learning capabilities, and built-in NLP features. It simplifies integrations with many other popular communication platforms. DialogFlow offers the ability to create highly intelligent chatbots that can understand natural language and keep improving over time.
Amazon Lex[10]	A part of Amazon Web Services (AWS) and is one of the most powerful and capable options available. It comes with built-in machine learning and NLU capabilities which make it highly scalable.

In this paper, we have used Xatkit Development Kit [1] to build the chatbot. Xatkit provides a set of Domain-Specific Languages to define chatbots in a platform-independent way. It comes with a Runtime engine that automates deployment of the chatbot application and manages the conversation over the platforms of choice. Fig.1 shows an overview of the Xatkit Framework.

2.3. Deep Learning and NLP

Currently, thanks to Deep Learning, Neural Networks have become the dominant approach for a wide variety of domains such as Image Recognition, Natural Language Processing and so on. For Natural Language Processing, Neural networks are used for Text Classification, Machine Translation, Named Entity Recognition, and many other tasks. In Conversational Agents systems, Neural Networks may be used for either end-to-end system training (Generative Chatbot models), or, in specific parts in the system.

2.4. Intent Recognition

Intent Detection, the main component of conversational systems, is considered as a classification task. The aim of this task is to classify the user utterance to a predefined class (intent), that words understand the user's goal. Before the spread of neural network techniques, Intent Detection was based on pattern-based recognition. Nowadays,

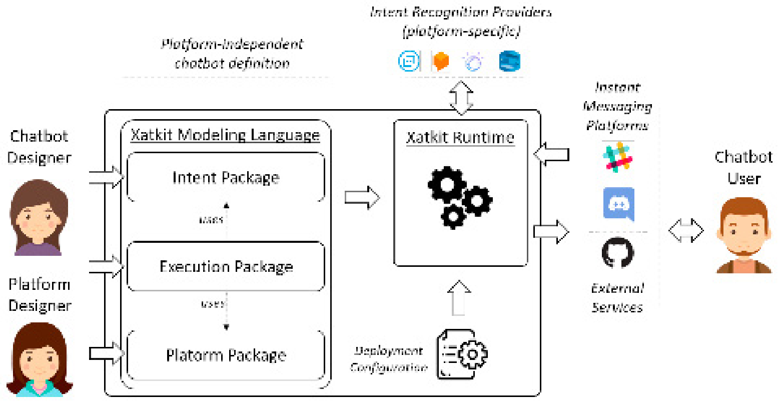

Fig. 1. Overview of the Xatkit Framework

neural networks have proven their effectiveness on this task. In the most recent studies, Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM) networks have been used in intent classification [11],[12]. However, this classification task suffers from a very small amount of data available to train the model and the utterances to be classified are generally short. Since Deep Learning models require a large amount of data, Pre-trained word vectors have been widely used in different NLP tasks and have proven good performance improvements. Recently, pre-trained language models, such as ULMFiT, ELMO, GPT, and BERT, have shown to be very effective for learning different languages by exploiting large amounts of unlabeled data. CamemBERT [2] is a state-of-the-art language model for French based on the RoBERTa model, a version of BERT, is pre-trained on a french dataset and evaluated on different NLP sub-tasks such as Part-Of-Speech (POS) tagging, Named Entity Recognition (NER) and others. It has shown effective results compared to previous monolingual and multilingual approaches.

3. Proposed System

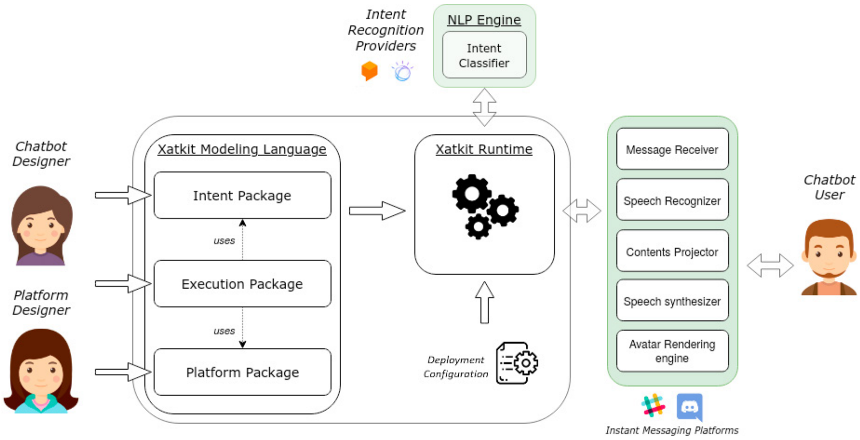

Fig. 2. Overview of our proposed system

In this paper, we propose a chatbot system that first receives the user's utterance (text or voice), recognizes the voice through the Speech Recognizer module, which transforms the voice input as a plain text. User input is received by the framework through Xatkit Runtime component, which is an event-based execution engine that deploys and manages the execution of the chatbot. The Runtime engine calls the Intent Recognition Provider module, which detects the user intention underlying the text input and consists of two components: CamemBERT encoder and intent classification decoder. Then, the resulting recognized intent is returned to the runtime, which performs a lookup and retrieves the list of actions and events associated with the recognized intent. Finally, those actions may be simple responses, which would be turned into speech by using Speech Synthesizer module or non-messaging actions such as showing animations, videos, starting a quiz and many more through a contents projector module. An animated 3D avatar will react with specific gestures, mimics and continues the conversation based on the given response with a synchronisation with the audio playback. In this section, we describe an overview of the proposed system. First, the Speech recognizer module is described. Then, the Encoder-Decoder model is detailed according to the Intent Recognition. Finally, Speech Synthesize and Avatar Rendering engine modules are detailed, respectively.

Fig.2 represents the architecture of the system.

3.1. Speech Recognizer

This module is responsible for processing incoming audio from messaging platforms and turning it into a plain text. There exist a couple of Speech Recognition providers which makes it easier for users to opt for one or another. In this work, we used Google's Speech-To-Text (STT) API, which is an online API that supports more than 125 languages, delivers important accuracy, and allows speech recognition in real-time.

3.2. Intent Recognition

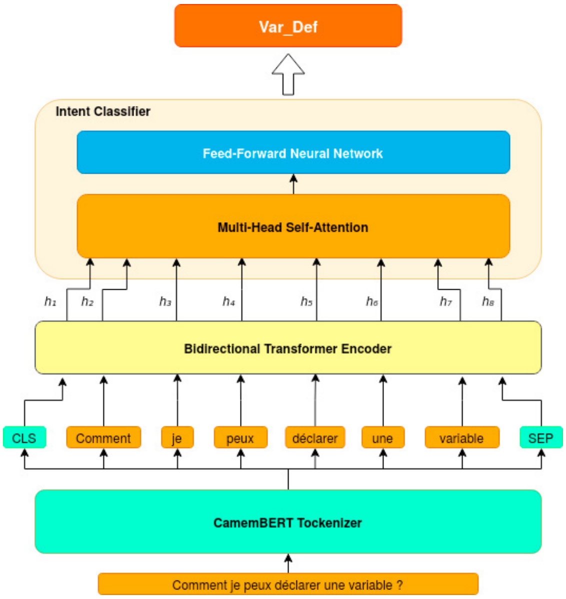
Var_Def
Intent Classifier
Feed-Forward Neural Network
Multi-Head Self-Attention
h₁
h₂
h₃
h₄
h₅
h₆
h₇
h₈
Bidirectional Transformer Encoder
CLS
Comment
je
peux
déclarer
une
variable
SEP
CamemBERT Tockenizer
Comment je peux déclarer une variable ?

Fig. 3. The overview architecture of CamemBERT for Intent Recognition. hi is each token's contextual semantic representation embedding.

CamemBERT encoder

The model is based on the CamemBERT model and represented in Fig.3. CamemBERT consists of several bidirectional transformer encoder layers that use the Multi-Headed Self-Attention mechanism, which learns contextual relations between tokens in text.

For sentence representation, CamemBERT takes the input utterance, as an input sequence. The input representation is the sum of the word embeddings and segment embeddings. In our case, utterance is, in general, a single sentence, making segment embeddings useless. A [CLS] token is added as the first token of the input sequence, and [SEP] token is inserted at its end. The input sequence is then forwarded to the CamemBERT encoder, which computes the contextual semantic representations of each token *hi*.

Intent classification decoder

Recently, the Self-Attention mechanism [13] was widely used in Machine Translation and Semantic Role Labeling tasks and achieved excellent results. Indeed, we exploit the strong modeling capacity of self-attention to learn the dependencies between tokens in a sequence. In our system, we adopt a multi-head attention mechanism, which improves the performance of the attention layer by encoding multiple relationships and nuances for each word. In fact, each head can focus on different projections of each word. For instance, one head could calculate the preposition/location relationships, while another head could calculate subject/verb relationships, simply by using different projections to create the Query, Key, and Value vectors, which are calculated by multiplying each output of the encoder by a set of different weight matrices W^Q, W^K and W^V to produce Q, K and V respectively. The outputs from each head are concatenated back in a large vector as follow:

$$\begin{aligned} &MultiHead(Q,K,V) = Concat(head_1,\ldots,head_h)W_o. \\ &\text{where} \quad head_i = Attention(QW_i^Q, KW_i^K, VW_i^V). \\ &\text{and} \quad W_o \quad \text{is trainable weight matrices} \end{aligned} \tag{1}$$

Finally, we predict the intent as follow:

$$Intent = softmax(MultiHead(Q,K,V)). \tag{2}$$

3.3. Speech Synthesizer

The next module is Speech Synthesizer. Like the Speech Recognizer module, there are a set of online and offline speech generators. Our system uses Google's Text-To-Speech (TTS), which is an online API. It delivers voices that are near human quality across more than 40 languages and variants.

3.4. Animated avatar

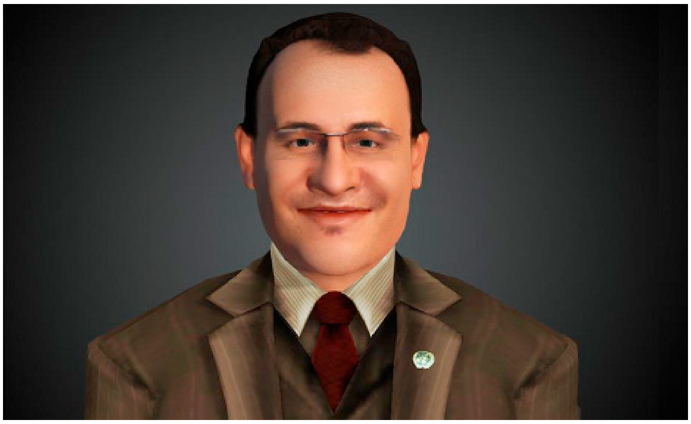

Fig. 4. Edu-Chatbot animate 3D-Avatar

The Final step is to build a digital human, which is an animated avatar that can produce a range of human body languages (gestures, mimics, etc). To achieve this task, we create a 3D avatar using Character Creator software, which is a character creation solution that eases creating, importing, and customizing realistic-looking character assets. The 3D avatar is animated through CrazyTalk software, which is a facial animation software that uses voice and text to vividly animate facial images, it provides interactions using verbal and non-verbal cues and comes with smooth lip-syncing results for any talking. All we need is to forward the speech to the CrazyTalk software and it will animate our 3D avatar to synchronize lips to the provided audio.

4. Conclusion

In this paper, we propose a chatbot system architecture using the Xatkit development framework. First, the system receives user's messages via an instant messaging platform. The received message can be a plain text or an audio. A speech recognizer module is used to transform audios to text. The recognized utterance is then forwarded to an in-house NLP Engine, which is based on the CamemBERT architecture, for Intent Recognition tasks through Xatkit's runtime engine. The recognized intent is returned to the runtime, and a lookup is performed to retrieve the list of actions and events associated with the recognized intent. Finally, the text response will be turned into speech through the Speech Synthesizer module and non-messaging actions are projected through a contents projector module. An animated 3D avatar will react with specific gestures, mimics and continues the conversation based on the given response with a synchronisation with the audio playback.

Our model will be evaluated then applied in Student Support scenarios. Indeed, as a part of our teaching initiative that aims to bring this model to the classroom as a tool to teach students, we plan to implement it in several institutions that allows us to conduct an initial validation of the usefulness and benefits of Edu-Chatbot in the classroom.

Acknowledgements

This work was supported by Al-Khawarizmi program for research support in the field of Artificial Intelligence and its applications (Morocco).

References

[1] G. Daniel, J. Cabot, L. Deruelle and M. Derras, ”Xatkit: A Multimodal Low-Code Chatbot Development Framework,” in IEEE Access, vol. 8, pp. 15332-15346, 2020, doi: 10.1109/ACCESS.2020.2966919.

[2] Martin, L. et al. CamemBERT: a Tasty French Language Model. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 7203–7219), 2020. Association for Computational Linguistics.

[3] D. A. Ferrucci, ”Introduction to “This is Watson”,” in IBM Journal of Research and Development, vol. 56, no. 3.4, pp. 1:1-1:15, May-June 2012, doi: 10.1147/JRD.2012.2184356.

[4] Michael Washington. 2016. An Introduction to the Microsoft Bot Framework: Create Facebook and Skype Chatbots using Microsoft Visual Studio and C # (1st. ed.). CreateSpace Independent Publishing Platform, North Charleston, SC, USA

[5] Williams, J. et al. “Fast and easy language understanding for dialog systems with Microsoft Language Understanding Intelligent Service (LUIS).” SIGDIAL Conference (2015).

[6] Hoy, M.. “Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants.” Medical Reference Services Quarterly 37 (2018): 81 - 88.

[7] Wit.AI. Available: https://wit.ai

[8] Bocklisch, Tom et al. “Rasa: Open Source Language Understanding and Dialogue Management.” ArXiv abs/1712.05181 (2017): n. pag.

[9] Google's DialogFlow. Available: https://dialogflow.com

[10] Amazon Web Services Inc.(2016) Amazon Lex: Conversational interfaces for your applications. [Online]. Available: https://aws.amazon.com/lex/

[11] Y. Kim, “Convolutional neural networks for sentence classification,” 2014, arXiv:1408.5882. Available: https://arxiv.org/abs/1408.5882

[12] S. Ravuri and A. Stolcke, “Recurrent neural network and LSTM models for lexical utterance classification,” in Proc. 16th Annu. Conf. Int. Speech Commun. Assoc., 2015, pp. 1–5.

[13] Ashish Vaswani, et al. 2017. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 6000–6010.