



BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS  
FACULTY OF ELECTRICAL ENGINEERING AND INFORMATICS  
DEPARTMENT OF AUTOMATION AND APPLIED INFORMATICS

# Deep Learning Based Chatbot Models

SCIENTIFIC STUDENTS' ASSOCIATIONS REPORT

Author:  
Richárd Krisztián Csáky

Supervised by  
Gábor Recski

2017

## Kivonat

A konverzációs ágens (chatbot) egy olyan program, mely természetes nyelvet használva képes emberekkel kommunikálni. A beszélgetés modellezése fontos feladat a természetes nyelvfeldolgozás és mesterséges intelligencia (MI) területén. Az MI tudományág megszületése óta egy jól működő chatbot létrehozása még mindig az egyik legnehezebb kihívás. A chatbotok sokféle feladatra használhatók, de mindegyik esetében elvárt, hogy megértsék a felhasználó mondandóját és az adott problémához releváns válaszokat generáljanak.

A múlt chatbot architektúrái kézi szabályokra és sablonokra, vagy egyszerű statisztikai módszerekre támaszkodtak. 2015 óta, a mélytanulás (deep learning) elterjedésével ezek a modellek gyorsan felcserélődtek elejétől végéig tanítható neurális hálózatokkal. Manapság a rekurrens enkóder-dekóder modell [Cho et al., 2014] dominál a konverzáció modellezésben. Ezt az architektúrát a neurális gépi fordítás területéről adaptálták, ahol rendkívül jó eredményeket ért el. Azóta sokféle változata [Serban et al., 2016] és kiegészítése született annak érdekében, hogy minél jobb minőségű legyen a chatbotok által folytatott beszélgetés.

Munkám során részletes irodalmi kutatást végeztem, melyben az elmúlt 3 évben publikált, több mint 70, a chatbotokkal kapcsolatos publikációt vizsgálok meg. Ezután amellet érvelek, hogy a konverzáció modellezés sajátosságai a jelenlegi state-of-the-art architektúráktól eltérő megközelítést igényelnek. Szakirodalmi példákra alapulva bemutatom, hogy a jelenlegi chatbot modellek miért nem vesznek figyelembe elég ún. prior a válasz generálása során, és ez hogyan befolyásolja a beszélgetés minőségét. Ezek a priorok olyan külső információt hordoznak, melyen a beszélgetés kondicionálva lehet, mint például a beszélők személye [Li et al., 2016] vagy hangulata. Amellet, hogy bemutatom az okait, javaslatokat is teszek a probléma orvoslására.

A dolgozat következő részében egy nemrég bemutatott modellt, mely jelenleg state-of-the-art-nak számít a neurális gépi fordításban, az úgynevezett Transformer-t [Vaswani et al., 2017] adaptálom a beszélgetés-modellezés feladatára. Először az eredeti cikkben leírt modell tanításával kísérletezek, tanítóadatként a Cornell Movie-Dialog Corpus [Danescu-Niculescu-Mizil and Lee, 2011] dialógusait használva. Emellet továbbfejlesztem a modellt saját, az enkóder-dekóder architektúra hiányainak orvoslására született ötletekkel. További priorokat adok bemenetként a modellbe, mint a beszélgetők személye vagy hangulata. Végül korábbi chatbot modellekkel való összehasonlítás útján részletes elemzést végzek arról, hogy az eredeti modell mennyire teljesít jól dialógus adattal és hogyan befolyásolják a generált válaszok minőségét az általam implementált további kiegészítések.

## Abstract

A conversational agent (chatbot) is a piece of software that is able to communicate with humans using natural language. Modelling conversation is an important task in natural language processing and artificial intelligence (AI). Indeed, ever since the birth of AI, creating a good chatbot remains one of the field's hardest challenges. While chatbots can be used for various tasks, in general they have to understand users' utterances and provide responses that are relevant to the problem at hand.

In the past, methods for constructing chatbot architectures have relied on hand-written rules and templates or simple statistical methods. With the rise of deep learning these models were quickly replaced by end-to-end trainable neural networks around 2015. More specifically, the recurrent encoder-decoder model [Cho et al., 2014] dominates the task of conversational modelling. This architecture was adapted from the neural machine translation domain, where it performs extremely well. Since then a multitude of variations [Serban et al., 2016] and features were presented that augment the quality of the conversation that chatbots are capable of.

In my work, I conduct an in-depth survey of recent literature, examining over 70 publications related to chatbots published in the last 3 years. Then I proceed to make the argument that the very nature of the general conversation domain demands approaches that are different from current state-of-the-art architectures. Based on several examples from the literature I show why current chatbot models fail to take into account enough priors when generating responses and how this affects the quality of the conversation. In the case of chatbots these priors can be outside sources of information that the conversation is conditioned on like the persona [Li et al., 2016] or mood of the conversers. In addition to presenting the reasons behind this problem, I propose several ideas on how it could be remedied.

The next section of my paper focuses on adapting the very recent Transformer [Vaswani et al., 2017] model to the chatbot domain, which is currently the state-of-the-art in neural machine translation. I first present my experiments with the vanilla model, using conversations extracted from the Cornell Movie-Dialog Corpus [Danescu-Niculescu-Mizil and Lee, 2011]. Secondly, I augment the model with some of my ideas regarding the issues of encoder-decoder architectures. More specifically, I feed additional features into the model like mood or persona together with the raw conversation data. Finally, I conduct a detailed analysis of how the vanilla model performs on conversational data by comparing it to previous chatbot models and how the additional features, affect the quality of the generated responses.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>History of Chatbots</b>	<b>5</b>
2.1	Modelling Conversations . . . . .	5
2.2	Early Approaches . . . . .	5
2.3	The Seq2seq Model . . . . .	5
<b>3</b>	<b>Background</b>	<b>6</b>
3.1	Recent Chatbot Architectures and Augmentations . . . . .	6
3.1.1	Attention . . . . .	6
3.1.2	Context . . . . .	6
3.1.3	Objective Functions . . . . .	6
3.1.4	Additional Features . . . . .	6
3.1.5	Knowledge Bases, Copying and Information Retrieval . . . . .	6
3.1.6	Task-Oriented Approaches . . . . .	6
3.1.7	Decoding and Beam Search . . . . .	6
3.1.8	Reinforcement Learning . . . . .	6
3.1.9	Pretraining . . . . .	6
3.1.10	Additional Encoder-Decoder Models . . . . .	6
3.1.11	Evaluation Methods . . . . .	6
3.2	Criticism . . . . .	6
3.2.1	Datasets . . . . .	6
3.2.2	Loss Function . . . . .	6
3.2.3	Memory . . . . .	6
3.2.4	Evaluation Metrics . . . . .	6
3.3	Summary . . . . .	6
<b>4</b>	<b>Experiments</b>	<b>7</b>
4.1	The Tranformer Model . . . . .	7
4.2	Datasets . . . . .	7
<b>5</b>	<b>Results</b>	<b>8</b>
<b>6</b>	<b>Future Work</b>	<b>9</b>
<b>7</b>	<b>Conclusion</b>	<b>10</b>
	<b>References</b>	<b>11</b>

# 1 Introduction

A conversational agent (chatbot) is a piece of software that is able to communicate with humans using natural language. Ever since the birth of AI, modelling conversations remains one of the field's toughest challenges. Even though they are far from perfect chatbots are now used in a plethora of applications like Apple's Siri [Apple, 2017], Google's Google Assistant [Google, 2017] or Microsoft's Cortana [Microsoft, 2017]. In order to fully understand the capabilities and limitations of current chatbot architectures and techniques I conduct an in-depth survey, where I examine related literature published over the past 3 years and I implement my own chatbot based on a novel neural network model.

The paper begins with a brief overview of the history of chatbots in Section 2, where I discuss the properties and objectives of conversational modelling. I present early approaches and the current dominating model based on neural networks for building conversational agents.

In the following section I present key architectures and techniques that were developed over the past 3 years relating to chatbots. I group publications into 11 groups based on the specific technique or approach that is discussed by the authors. After this I present criticism regarding some of the properties of current chatbot models and I show how several of the techniques used are inappropriate for the task of modelling conversations.

In the next section I conduct experiments by training a novel neural network model, the Transformer [Vaswani et al., 2017] using dialog datasets [Danescu-Niculescu-Mizil and Lee, 2011, Tiedemann, 2009]. I run several trainings using these datasets detailed in Section 4. I present the results of the different training setups by qualitatively comparing them to previous chatbot models and by using standard evaluation metrics.

In the final section before concluding I offer possible directions for future work. More specifically, I propose several ideas in order to remedy the problems discussed in Section 3.2.

## **2 History of Chatbots**

### **2.1 Modelling Conversations**

Chatbot models usually take in as input natural language sentences uttered by the user, and output a response. Models have also been developed to accommodate for spoken or visual inputs. They oftentimes make use of a speech recognition component to transform speech into text [Serban et al., 2017] or convolutional neural networks that transform the input pictures into useful representations for the chatbot. The latter models are also called visual dialog agents, where the conversation is grounded on both textual and visual input [Das et al., 2017].

Mik azok a chatbotok Miben mas mint a task-specific dolgok ...

### **2.2 Early Approaches**

### **2.3 The Seq2seq Model**

## **3 Background**

### **3.1 Recent Chatbot Architectures and Augmentations**

#### **3.1.1 Attention**

#### **3.1.2 Context**

#### **3.1.3 Objective Functions**

#### **3.1.4 Additional Features**

#### **3.1.5 Knowledge Bases, Copying and Information Retrieval**

#### **3.1.6 Task-Oriented Approaches**

#### **3.1.7 Decoding and Beam Search**

#### **3.1.8 Reinforcement Learning**

#### **3.1.9 Pretraining**

#### **3.1.10 Additional Encoder-Decoder Models**

#### **3.1.11 Evaluation Methods**

### **3.2 Criticism**

#### **3.2.1 Datasets**

#### **3.2.2 Loss Function**

#### **3.2.3 Memory**

#### **3.2.4 Evaluation Metrics**

### **3.3 Summary**

## **4 Experiments**

### **4.1 The Transformer Model**

### **4.2 Datasets**



## 5 Results

## **6 Future Work**

## **7 Conclusion**

## References

- [Apple, 2017] Apple (2017). Siri. <https://www.apple.com/ios/siri/>. Accessed: 2017-10-04.
- [Cho et al., 2014] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
- [Danescu-Niculescu-Mizil and Lee, 2011] Danescu-Niculescu-Mizil, C. and Lee, L. (2011). Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, pages 76–87. Association for Computational Linguistics.
- [Das et al., 2017] Das, A., Kottur, S., Moura, J. M., Lee, S., and Batra, D. (2017). Learning cooperative visual dialog agents with deep reinforcement learning. *arXiv preprint arXiv:1703.06585*.
- [Google, 2017] Google (2017). Google assistant. <https://assistant.google.com/>. Accessed: 2017-10-04.
- [Li et al., 2016] Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., and Dolan, B. (2016). A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155*.
- [Microsoft, 2017] Microsoft (2017). Cortana. <https://www.microsoft.com/en-us/windows/cortana>. Accessed: 2017-10-04.
- [Serban et al., 2017] Serban, I. V., Sankar, C., Germain, M., Zhang, S., Lin, Z., Subramanian, S., Kim, T., Pieper, M., Chandar, S., Ke, N. R., et al. (2017). A deep reinforcement learning chatbot. *arXiv preprint arXiv:1709.02349*.
- [Serban et al., 2016] Serban, I. V., Sordoni, A., Bengio, Y., Courville, A. C., and Pineau, J. (2016). Building end-to-end dialogue systems using generative hierarchical neural network models. In *AAAI*, pages 3776–3784.
- [Tiedemann, 2009] Tiedemann, J. (2009). News from opus-a collection of multilingual parallel corpora with tools and interfaces. In *Recent advances in natural language processing*, volume 5, pages 237–248.
- [Vaswani et al., 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*.