

REVIEW

by Prof. Dr.Sc. Galia Angelova, Department of AI and Language Technologies,
Institute of Information and Communication Technologies, Bulgarian Academy of Sciences

for Momchil Emilov Hardalov's thesis

"Intelligent Context-Aware Natural Language Dialogue Agent"

submitted in fulfillment of the requirements of the educational and scientific degree "Doctor
of Philosophy" (PhD) in Professional field: 4.6 Informatics and Computer Science,

Doctoral program: "Software Technologies" – Knowledge Discovery

Pursuant to order No. RD-38-600/28.10.2022 of the Rector of Sofia University "St. Kliment Ohridski" (SU) I participate in the Scientific Jury for awarding the educational and scientific degree "PhD" to Momchil Emilov Hardalov, doctoral student at the Faculty of Mathematics and Informatics (FMI) in 2017-2021 who completed his doctoral studies with right of defense. The topic of the thesis is focused on the challenges of creating high-quality "conversational agents" for the needs of practical applications: trainings, explanations to customers, question answering, advisory dialogue systems, and virtual assistants. This topic is among the hottest trends in man-machine interfaces, which are constantly expanding our understanding of human-computer interaction. Interface automation and agent reliability are priorities in the development of high quality customer services. So the first positive impression of the dissertation comes from the ambitious choice of an advanced field in which applications will be developed over a relatively long period in the future.

Formally, Art. 26, para. 2 and 3 of the Act on Development of Academic Staff in the Republic of Bulgaria requires candidates to obtain the PhD degree in specialty 4.6 "Informatics and computer sciences" to meet the minimal requirements of 50 points for indicator A (presentation of doctoral thesis) and 30 points for indicators group "Г". Momchil Hardalov declares the fulfillment of the minimal national requirements in specialty 4.6 by presenting a list of four scientific papers in English where he is a co-author. These papers were published in the period 2018-2020, they present the results of the dissertation thesis, all are indexed by Scopus and three of them have an SJR-rank. The points collected from this publication activity (more than 108) exceed three times the requirements for a minimum threshold of 30 points in the group "Г" indicators. With this, the formal requirements for satisfying the national criteria for the defense of a doctoral dissertation have been met and the defense procedure of the presented dissertation can be started. In the Chapter 1, two more papers are listed that present thesis results. A total of 77 citations of papers related to the dissertation were found. When searching Google Scholar on 14/01/2023 for M. Hardalov, 20 articles in computational linguistics appear, all in English and in prestigious international proceedings, with a total of 413 citations (self-citations are included, but nevertheless the number is impressive). This profile displays Mr. Hardalov as a mature expert in natural language processing.

The objectives of the thesis are (i) to create efficient approaches for building context-aware conversational agents using free text for specific tasks such as customer service chatbots, and (ii) to create new language resources supporting the development of an intelligent agent for multilingual dialogue that can generate long answers (e.g. articles) as opposed to the usual short answers. The thesis tasks are to design, develop and evaluate (i) a natural language understanding component that identifies user intent and recognises what is relevant to its slots; (ii) an utterance synthesis algorithm based on external knowledge sources; (iii) end-to-end generative models for chatbots with application in customer support, conducting dialogue with non-encountered sentences and (iv) a system for multilingual and cross-lingual dialogue for application in a conversational agent.

The thesis contains 154 pages and two appendices A and B of 8 pages. It is structured in 6 Chapters and includes 18 figures, 39 tables, a bibliography of 350 titles in English, acknowledgments and a list of abbreviations.

Thesis Content and Contributions

Chapter 1 (Introduction) convincingly motivates the relevance and importance of the research and introduces the main concept: "*A conversational agent also referred to as chatbot is a computer program which tries to generate human like responses during a conversation*". The agent has at least three functionalities: *question answering* (synthesizes short, direct answers to user queries using rich information extracted from various sources); *completing tasks* such as restaurant reservations, scheduling meetings; and *social chat*: to converse seamlessly and appropriately with users and provide useful recommendations.

Chapter 2 (Background and Related Work) is a compact and concise summary of current approaches to creating conversational agents. The chapter examines the individual agent components: understanding the topic (intent) of the request and identifying relevant slots in the dialog, extracting the response with a focus on detecting already verified claims, as well as techniques for dialog implementation and combining responses. The linguistic resources needed to build conversational agents are also discussed.

Chapters 3, 4 and 5 respectively present the results achieved in solving the main tasks of the thesis ‘to do’ list: understanding the user intent in the free text input, finding the answer, and carrying out the dialogue with the user. The solutions presented are not unified into a single conversational agent prototype (which can be expected in principle due to the thesis title – “an agent” in singular). Chapter 6 contains the conclusion, a statement of contributions and plans for future work.

Chapter 3 (Semantic Parsing of Human-Generated Utterances) presents an enriched neural model Transformer-NLU for simultaneous intent classification and slot filling, which uses the pre-trained BERT and RoBERTa. The model is tested for English over the corpora ATIS (aircraft flight information requests) and SNIPS (collected by a custom intent engine for

personal voice assistants). An attention pooling layer is introduced for intent classification, which uses a holistic representation of the input sentence. The improvements are in three directions: for intent detection, a unified representation from the last hidden states for all tokens is used; the model distinguishes the named units thanks to the capital letters (additional features); and the slot filling layer is fed with the predicted intent distribution, the last hidden BERT's representations and the word features. Testing was performed removing and adding the new components to evaluate the contribution of the proposed improvements to the results. Compared with previous developments, Transformer-NLU shows better results than other models based on multilayer neural networks.

Chapter 4 (Curating answers from external knowledge sources) presents results related to finding the answer to the received user request. Retrieving already published articles or other ready-made documents that contain information about the correct answer provides the agent with the ability to generate already verified statements. Two scenarios are explored: first, answering multiple-choice questions using extracted contexts with evidence paragraphs from an external knowledge base, and second, deriving an answer from a collection of long explanations (documents or articles). In the first scenario, neural models that use external knowledge in a multilingual environment of languages with rich resources (English) and few resources (Bulgarian) are considered. Most convenient for experimentation are corpora of question-answer pairs, for which appropriate explanatory passages are sought. In addition to the RACE corpus with over 87,000 school-level multiple-choice questions in English, a new Bulgarian language corpus with over 2,600 questions has been created. A multilingual BERT model pre-trained over English or SlavicBERT over Slavic languages (Bulgarian, Czech and Polish Wikipedia articles as well as Russian news) were used. For Bulgarian, the best accuracy of over 42% was obtained. It is interesting to observe that prior training on Slavic languages causes "catastrophic forgetting" of learned English syntax and semantics (and the reverse is not true). In the second scenario, answer retrieval is interwoven with fact-checking: for a given user comment, check whether the stated claim has already been verified. The dissertation examined tweets from the Snopes website for the period October 2017 - October 2021. A new large CrowdChecked corpus was created with nearly 150,000 unique pairs <tweet> - <retweet of "fact-checker from the crowd" with a link to proof article> and more than 200k other tweets included in the conversations. Considerable work has been invested into manually parsing and structuring the conversations (assuming that the "man in the crowd" is error-free). The CrowdChecked corpus is used for fact-checking experiments, where the input data are the pairs <tweet> - <fact-checking article> labeled in two ways: according to Jaccard similarity and the predictions of a multilayer neural network. The base model is Sentence-BERT (SBERT) with extensions to independently assess lexical and semantic similarities, an optimized temperature parameter in the error function, learning with noisy data using self-adaptive learning, and additional weights in the error function. The model achieves a mean average precision MAP above 90%, better than other models in the CLEF 2021 competition.

Chapter 5 (Advanced Conversation) presents results in the area of dialog management and generation of previously unused new agent utterances without relying on external sources. Automated customer service prototypes are trained on corpora of tweets and responses from

real company conversations conducted in customer dialogue. Experiments have been done with Twitter conversations with two types of agents: those that extract information or that use neural networks to synthesize answers to questions. In the evaluation, carried out over about 50 thousand question-answer pairs, metrics assessing word overlap and semantics were used. The better quality of the answers synthesized by the generative models based on neural networks has been shown, because they manage to generate grammatically correct and appropriate answers to the questions they are asked. On the other hand, with insufficient sample questions and answers for training on a particular topic, agents do not perform well. An approach is proposed for ranking candidates for next answer, which involves training a classifier to evaluate the quality of a question-answer pair. This chapter also introduces a new multilingual Exams corpus of questions and answers in 16 languages at school level in 24 subjects, which can be used in multilingual scenarios for training question-answer systems. The corpus is built primarily on encyclopedic (verified) facts from Wikipedia and can be used to develop multilingual models that can reason with learned knowledge. As it is mentioned in the indicated topics for future work, Exams may (according to the author) be used as a factual information base that can become a basis for training intelligent models capable of explaining knowledge stored in the model.

Chapter 6 (Conclusion and Future Work) contains a summary of the results and lists the original research and applied research contributions of the dissertation. Directions for future work include plans to improve the models (to increase their capacity, ensure efficient updating, protect against maliciousness, achieve explainability, etc.). This chapter lacks an "approbation" section to list public presentations of dissertation results at conferences and seminars.

Discussion

The dissertation presents technological results obtained with extended models and enriched techniques that are based on state-of-the-art language representation models. A variety of new language resources have also been created to experiment with the proposed extended models, surprising the reader with the effort involved in collecting corpora of tweets or school tests. Due to the concise presentation of a large volume of technical information, a number of questions remain unanswered, for example, the relatively simple example of a slot in the article "Enriched Pre-trained Transformers for Joint Slot Filling and Intent Detection" raises the question of the scalability of the approach to understanding more complex requests/intents and slots. Another question in this chapter concerns the role of capital letters in the recognition of named entities, bearing in mind that they can be placed incorrectly when recording by voice assistants. There are also questions on the generated responses shown, e.g. why only the IR-model in Tabl. 5.5 on page 82 uses "please"? In general, it may be recommended to include in the thesis more textual examples of the solutions proposed for the tasks discussed in Chapters 3, 4, and 5, or a separate thesis appendix with examples of understood intentions and slots, elicited responses, and generated utterances in the dialogues.

I accept the author's research and applied research contributions as declared in the Chapter 6 of the thesis. In my opinion, the most important achievement is that Momchil Hardalov got involved into the "hot" issue of using neural models for free text processing as well as his inclusion in the world practice of participation in international conferences, international author's collectives and established thematic competitions.

The text of the dissertation is concise and specific, appropriately organized with a clear and logical division into chapters and with references in each chapter to a list of author's publications in which the results are presented. The abstract correctly reflects the content of the thesis.

I accept that Mr. Hardalov is an equal co-author of the publications presenting the results of the doctoral thesis. Undoubtedly, he invested a very large share of the effort in the construction of the newly created language resources.

I do not detect plagiarism in the dissertation work, which is based on the texts of scientific publications of the listed co-authors.

I know Mr. Hardalov from his student years in the master's program "Information Extraction and Knowledge Processing" at the "Software Technologies" Department. It is pleasant to follow the growth of a young specialist in computational linguistics, even more so with a taste for processing Bulgarian text and creating language resources for Slavic languages.

Conclusion

The Act on the Development of Academic Staff in the Republic of Bulgaria requires that a doctoral thesis contains research or applied research results that represent an original contribution to science. The dissertation must demonstrate that the candidate has in-depth knowledge of the relevant specialty as well as ability for independent research. These requirements have been met for Momchil Hardalov's thesis and the scientific articles where the obtained results are tested. The large number of citations of the publications related to the dissertation, as well as of the candidate's other scientific papers, prove the wide international response and interest. On these grounds, I will vote positively for awarding the degree and I confidently suggest to the respected Scientific Jury to award Momchil Emilov Hardalov the educational and scientific degree "Doctor in Philosophy" (PhD) in professional field 4.6 "Informatics and Computer Sciences".

January 17, 2023

Member of the Scientific Jury:

Prof. Dr.Sc. Galia Angelova