

Annobot: Platform for Annotating and Creating Datasets through Conversation with a Chatbot

Rafał Poświata and Michał Perełkiewicz

National Information Processing Institute
al. Niepodległości 188b, 00-608 Warsaw, Poland
{rposwiata, mperelkiewicz}@opi.org.pl

Abstract

In this paper, we introduce Annobot: a platform for annotating and creating datasets through conversation with a chatbot. This natural form of interaction has allowed us to create a more accessible and flexible interface, especially for mobile devices. Our solution has a wide range of applications such as data labelling for binary, multi-class/label classification tasks, preparing data for regression problems, or creating sets for issues such as machine translation, question answering or text summarization. Additional features include pre-annotation, active sampling, online learning and real-time inter-annotator agreement. The system is integrated with the popular messaging platform: Facebook Messenger. Usability experiment showed the advantages of the proposed platform compared to other labelling tools. The source code of Annobot is available under the GNU LGPL license at <https://github.com/rafalposwiata/annobot>.

1 Introduction

The basis of any machine learning model is data. In the case of supervised solutions, such data must be labelled, very often manually, especially if problems are closely related to human interpretation such as emotion classification (Mohammad et al., 2018) or hate/offensive language detection (Zampieri et al., 2019). The preparation of such data sets is most often carried out using the crowdsourcing platforms such as Figure Eight (earlier called CrowdFlower)¹ or Amazon Mechanical Turk². These platforms undoubtedly have many advantages, but they are paid solutions that not everyone can afford. Furthermore, due to their commercial nature, no changes or modifications can be made on your own. There are also open-source solutions on the market, but you have to organize a group of labellers yourself. Another labelling issue is the lack of an intuitive, flexible, well-known, and responsive user interface. Taking into account the above and the fact that chatbot technology has recently become more widespread, we decided to create the Annobot platform. This is the first open-sourced platform for annotating and creating datasets through conversation with a chatbot. This natural form of interaction provides an interface that meets all criteria, while integration with the messaging application allows for more efficient activation of users.

The rest of the paper is organized as follows. Section 2 briefly describes open-source text annotation systems developed so far. Section 3 shows an overview of our platform and its detail functions. Experiment for showing the effectiveness of the created system is presented in Section 4. Next section presents possible applications and impact. Finally, Section 6 concludes this paper.

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>.

¹<https://www.figure-eight.com/>

²<https://www.mturk.com/>

2 Related Work

If we trace the history of labelling tools³, most of them have been created for sequence tagging (Stenetorp et al., 2012; Yimam et al., 2013; Bontcheva et al., 2013; Yang et al., 2018; Kummerfeld, 2019; Lin et al., 2019). Recent years have brought tools for other purposes such as annotating dialogues (Collins et al., 2019), documents (Nakayama et al., 2018), and other types of data (Heartex, 2019). When considering the user interface, most tools offer GUI (all mentioned before), although there are also alternative approaches such as command line (Yang et al., 2018; Kummerfeld, 2019). What’s worth mentioning is that the latest tools claim to support mobile versions (Nakayama et al., 2018; Lin et al., 2019; Heartex, 2019) and have more and more innovative functionalities, here we can mention AlpacaTag, which has, e.g. active, intelligent recommendations or automatic crowd consolidation module. Our solution is in line with current trends related to support for mobile devices and advanced additional functionalities (e.g. active sampling). However, we have applied a different form of interaction than in other tools, namely CUI (Conversational User Interface). Our tool is integrated with a platform with billions of users (Facebook) and has a very wide list of possible applications resulting from the CUI.

3 Platform Overview

As shown in Figure 1, the Annobot platform consists of five modules. **Annobot chat** and **Facebook Messenger** are dedicated to users/labellers, while **Annobot admin panel** is used to manage the platform (including adding data sets or preparing a conversation scenario). These are web-based components. The main module is the **Annobot core**. It is a server application, responsible for receiving, sending, and recording all operations performed within the platform and integrating it with Facebook. The **ML models** module is a set of machine learning models used for more advance functions described later on.

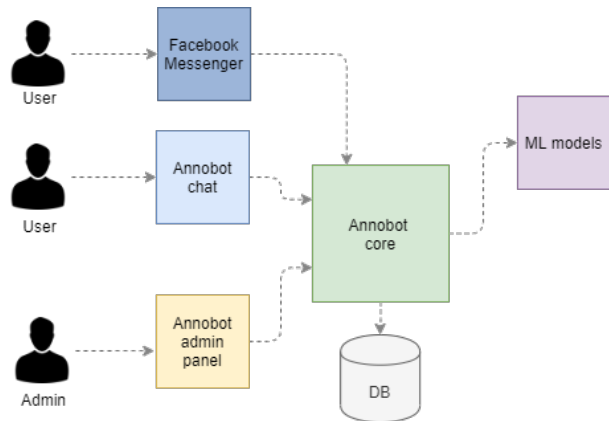
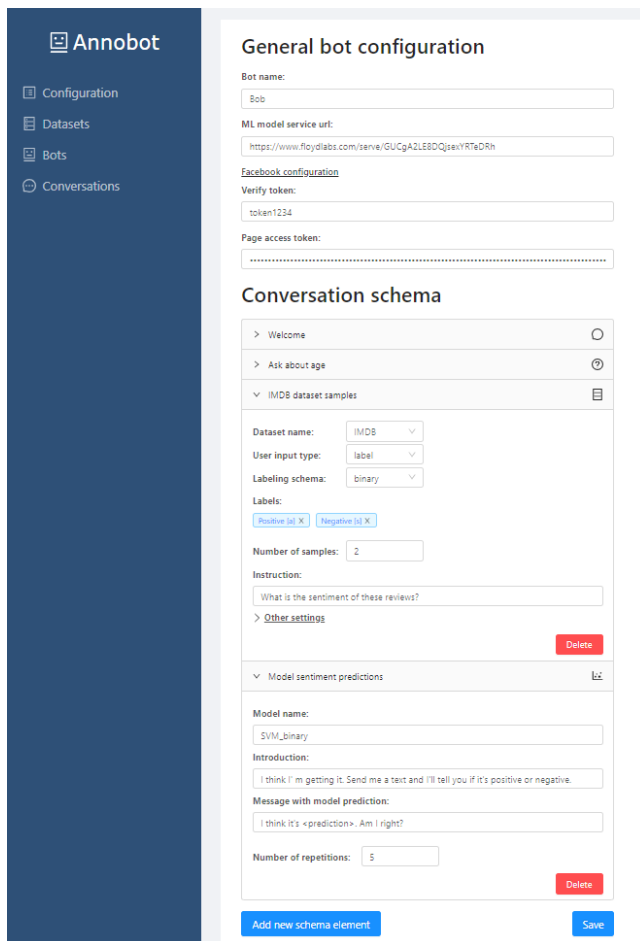


Figure 1: Overview of the Annobot platform.

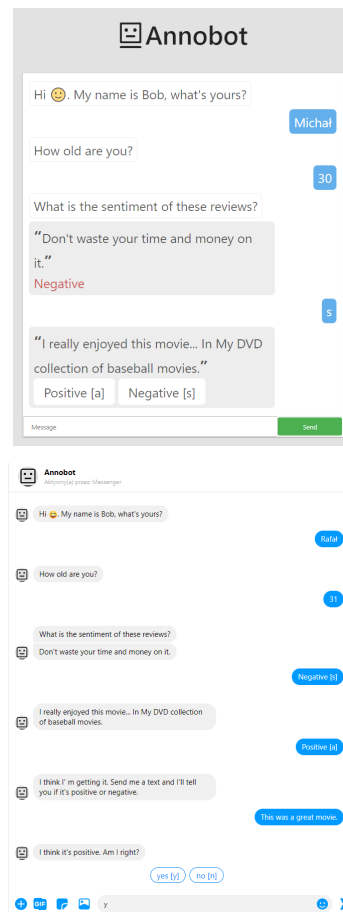
3.1 Functionalities

Annotating through conversation is the main functionality of our platform. To make a chatbot available to people, you must first configure it. We start by naming the bot and, if we want to use advanced features, adding the URL to the ml models. Then we can integrate our bot with Facebook. Next, you need to create a conversation schema. The schema is our planned scenario, according to which the bot works. It consists of any number of steps, belonging to one of four classes: simple message, question, sample for labelling or model prediction. In Figure 2a there is an example of such a configuration. In this example, the first two steps are a simple welcome message and the question about age. The third step corresponds to the actual labelling data. The administrator has to specify the data set, user input type (e.g. label), the labelling scheme (e.g., binary), possible labels (e.g., positive/negative), and instructions for the labellers. The last step allows you to test the chatbot (exactly saying the selected model - SVM.binary), to which the user can send his text. Examples of conversations between a labeller and a chatbot created according to this scheme is shown in Figure 2b. In addition to this functionality, the system also has other features, which we will present below.

³We only considered non-commercial tools.



(a) Chatbot configuration.



(b) Conversations between user and chatbot using Annobot chat and Facebook Messenger.

Figure 2: Configuration and conversations with chatbot.

3.1.1 Pre-annotation

Pre-annotation is a procedure of automatic annotation of the text using the existing automatic system and presenting these annotations to the human annotator. In our system, this functionality is realized, e.g. in the form of a message sent by a chatbot to a user, in which it asks about the correctness of the classified text. The user can then confirm or deny it.

3.1.2 Active Sampling

Sampling labelling data can affect the time spent on creating a valuable collection. In our platform, we use the ML models module to implement Least Confidence (Culotta and McCallum, 2005) active learning method. Firstly, the application sends examples with labels to it (one by one or in batches), and secondly, it uses its predictions to sort the unlabeled data.

3.1.3 Online learning

Online learning consists of continuous updating and improving the existing system/model. Data can be transferred after each assigned label or in batches. Annobot can be integrated with any model that has the appropriate REST API.

3.1.4 Inter-annotator agreement

Inter-annotator agreement (IAA) is a measure of how well two (or more) annotators can make the same annotation decision. It is an important part of any labelling tool used by several people in parallel.

3.2 Technology

When creating our platform, we tried to choose technologies that are well established and currently actively develop. To create web modules (chat and admin panel), we used HTML, CSS and TypeScript with ReactJs library. The core module was written in Java with the use of the Spring framework. The ML model was created in Python using libraries, such as scikit-learn, Flair and Flask. While as our data source, we used PostgreSQL database.

4 Usability Experiment

To verify the efficiency of our platform, we conduct a preliminary annotation experiment. For comparison with our system, we have chosen two tools that can be used for document annotating: Label Studio (Hearstex, 2019) and Doccano (Nakayama et al., 2018). We extracted 48 reviews from the IMDB dataset (Maas et al., 2011) as the corpus to be annotated. The task was to determine the sentiment (positive or negative) of the review. The job should be performed using each system first in the desktop version and then in the mobile version using a mobile device. Therefore the experiment consisted of 6 sub-tasks (3 tools x 2 versions). To each of this sub-task, we assigned eight reviews, so that they had similar text length distribution and the same class representation. The experiment was attended by 12 people (5 women and 7 men). To eliminate the “first system“ effect, participants were given instructions in which the order of the systems was different. The results are shown in Figure 3. When analyzing the average results for desktop versions, we found an only small difference between our solution and Label Studio according to the t-test with 0.05 significance level. However, the same test showed a significant difference between these systems and Doccano. Analyzing the results of the mobile version, we found that the difference between Annobot and either Label Studio or Doccano is significant at the 0.05 level, according to a t-test. In the case of mobile versions, the participants, speaking about the Annobot, unanimously pointed to intuitiveness and the form of interaction they are familiar with (they usually pointed out that they are Facebook Messenger users).

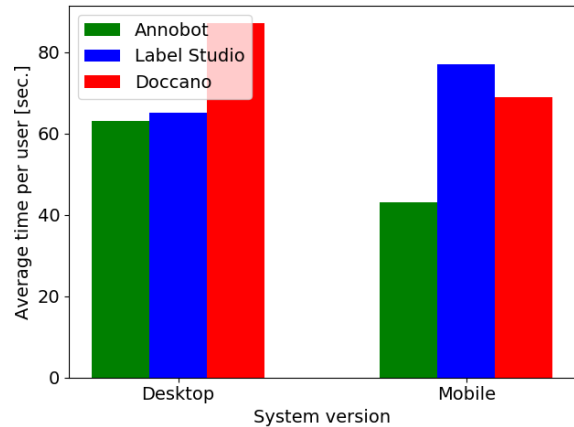


Figure 3: Average labelling time (in seconds) using Annobot, Label Studio, and Doccano in desktop and mobile versions.

5 Applications

The Annobot has many applications, such as data labelling for binary, multi-class/label classification tasks, preparing data for regression problems, or creating sets for issues such as machine translation, question answering or text summarization. If we are talking about the type of data: these can be short texts, sentences or tweets, as well as longer documents. Potential recipients of the platform may be researchers (not only in the field of computer science, but, e.g. sociologists or psychologists who can create surveys with the help of Annobot), companies that want to accelerate the development and improve their systems/models by making them accessible to the world, or ordinary people for whom teaching chatbot, e.g. to recognize what hate speech is, can be a form of participation in the development of AI, which has benefits for society.

6 Conclusion

In this paper, we propose the Annobot platform. We believe that our solution offers a lot of possibilities resulting from the form of interaction we have adopted. It is possible, for example, to introduce an element of “curiosity“ of the bot, e.g., by asking about the reason for this and not another decision of the labeller. In the future, we would like to integrate our platform with Slack⁴ communication platform.

References

- Kalina Bontcheva, Hamish Cunningham, Ian Roberts, Angus Roberts, Valentin Tablan, Niraj Aswani, and Genevieve Gorrell. 2013. Gate teamware: a web-based, collaborative text annotation framework. *Language Resources and Evaluation*, 47(4):1007–1029, Dec.
- Edward Collins, Nikolai Rozanov, and Bingbing Zhang. 2019. LIDA: Lightweight interactive dialogue annotator. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 121–126, Hong Kong, China, November.
- Aron Culotta and Andrew McCallum. 2005. Reducing labeling effort for structured prediction tasks. volume 2, pages 746–751, 01.
- Heartex. 2019. Label Studio. Software available from <https://github.com/heartexlabs/label-studio>.
- Jonathan K. Kummerfeld. 2019. SLATE: A super-lightweight annotation tool for experts. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 7–12, Florence, Italy, July.
- Bill Yuchen Lin, Dong-Ho Lee, Frank F. Xu, Ouyu Lan, and Xiang Ren. 2019. AlpacaTag: An active learning-based crowd annotation framework for sequence tagging. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 58–63, Florence, Italy, July.
- Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150, Portland, Oregon, USA, June. Association for Computational Linguistics.
- Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. SemEval-2018 task 1: Affect in tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 1–17, New Orleans, Louisiana, June.
- Hiroki Nakayama, Takahiro Kubo, Junya Kamura, Yasufumi Taniguchi, and Xu Liang. 2018. doccano: Text annotation tool for human. Software available from <https://github.com/doccano/doccano>.
- Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun’ichi Tsujii. 2012. brat: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107, Avignon, France, April.
- Jie Yang, Yue Zhang, Linwei Li, and Xingxuan Li. 2018. YEDDA: A lightweight collaborative text span annotation tool. In *Proceedings of ACL 2018, System Demonstrations*, pages 31–36, Melbourne, Australia, July.
- Seid Muhie Yimam, Iryna Gurevych, Richard Eckart de Castilho, and Chris Biemann. 2013. WebAnno: A flexible, web-based and visually supported system for distributed annotations. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 1–6, Sofia, Bulgaria, August.
- Marcos Zampieri, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, and Ritesh Kumar. 2019. Semeval-2019 task 6: Identifying and categorizing offensive language in social media (offenseval). In *SemEval@NAACL-HLT*.

⁴<https://slack.com/>