Intent Classification: French Recruitment Chatbot Use Case

Nadira Boudjani, Viviane Colas, Azade Fotouhi
SogetiLabs, Paris, France
{nadira.boudjani, viviane.colas, azade.fotouhi}@sogeti.com

***Abstract*—Intent classification is an important task for the chatbot to understand and interpret the user's message. For the proper functioning of the chatbot, it's essential to choose an effective intent classifier. In this paper, we propose a comparative study between three intent classifier models for a French recruitment chatbot; DIETClassifier (Rasa), Wit.ai Classifier and CamemBERT Classifier. First, We constructed a French dataset with intents in a recruitment chatbot context and trained all three classifiers on this dataset. According to our results CamemBERT is the best intent classifier for our chatbot.**

***Index Terms*—Intent classification, CamemBERT, Rasa, Wit.ai, Chatbot.**

I. Introduction

In recent years, the application of AI in the creation of chatbots has been significantly increased in various domains: education [1], [2], medicine [3], [4], recruitment [5]–[8] and multi-domain conversation generator such as ChatGPT [9].

A chatbot is a software program able to talk with a user, to understand the requests received and to reply to them. Intent classification is a crucial step for the proper functioning of the chatbot. It allows the user's will to be detected in the message. Figure 1 shows a few examples of user intent during a conversation with a chatbot.

Message	Intent
I would like to book a train ticket	book_train_ticket
I need more information about my last order	order_traking

Fig. 1. Intent classification examples.

Several frameworks have been developed for the creation of AI chatbots. Among the most popular: Rasa and Wit.ai.

- Rasa [10] is an open-source conversational AI platform for understanding and holding conversations, and connecting to messaging channels and third-party systems through a set of APIs.
- wit.ai [11] is a platform that provides services to create a dialoger through an open and extensible automatic natural language processing engine. It provides an easy interface and fast-learning APIs to understand human communication from every interaction.

These platforms are used in the development of several chatbots in different languages [12]–[14]. To the best of our knowledge, no comparison of these frameworks in intent classification for a French has been presented in the literature.

The only comparison for French is proposed in [15]. The authors present a comparison between two French versions of BERT, CamemBERT and FlauBERT, for a French medical chatbot.

Our aim is to test the two platforms above in the intent classification of a recruitment chatbot in French language and to compare these results with the CamemBERT Classifier results.

The rest of the paper is organised as follows: Section 2 presents chatbot frameworks, Section 3 presents our dataset, Section 4 presents results of the experiments. Finally, Section 5 presents the conclusion and the perspective.

II. Chatbot frameworks

In this section, we present Rasa and wit.ai components and how these frameworks perform intent classification.

A. Rasa framework

Rasa consists of 5 files :

- Config: Describes the language environment and adds functionality to Rasa such as Spacy, in order to manage French language for example.
- Domain: defines the environment in which the chatbot operates. It specifies the intents, entities, responses, and actions the chatbot should perform.
- NLU (Natural Language Understanding): performs intent classification, entity extraction, and response retrieval.
- Stories: Conversation stages based on exchanges between chatbot and human. It's a set of short scenarios.
- Rules: List of rules that the chatbot follows systematically, based on identified keywords.

Intent classification in Rasa is handled by DIETClassifier, a multi-tasking transformer architecture that enables both intent classification and entity recognition.

To train DIETClassifier in Rasa, we need to modify 3 default files:

- Config: we add the line "- name: DIETClassifier" and we can add other parameters such as the number of epochs (element of iterative process).
- Domain: the list of intents are given in Domain file preceded by the keyword "intents:" as shown in figure 2.
- NLU: in this file, for each intent, examples of sentences expressing this intent are given as shown in figure 3.

2769-5654/23/$31.00 ©2023 IEEE
DOI 10.1109/CSCI62032.2023.00117

2023 International Conference on Computational Science and Computational Intelligence (CSCI) | 979-8-3503-6151-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/CSCI62032.2023.00117

```
intents:
  - accueil
  - aurevoir
  - information
  - affirmation
  - erreur_utilisateur
  - refuser_repondre
  - negation
```

Fig. 2. Intents definition in the domain file example.

```
version: '3.0'
nlu:
- intent: accueil
  examples: |
    - Bonjour
    - Bonjour, comment-allez vous ?
    - Salut
    - Salut, ça va ?
    - Bonjour, comment vas tu ?
    - ça va
    - Bien et vous ?
```

Fig. 3. Example of sentences with greeting intention in NLU file.

B. Wit.ai framework

Wit.ai is organised into 3 parts: Composer, responsible for creating the chatbot; Management, where intentions and entities are defined; Understanding, responsible for training wit.ai models through examples.

The Composer has 4 components. Figure 4 shows part of the chatbot composer:

- Input: used to represent a message sent by a user.
- Response: used to reply to a user.
- Decision: used to direct the flow of information on the basis of conditions previously defined.
- Context: used to add, delete or modify information in the context (a JSON object) of the information flow during dialog.

Intents are defined in the Management (intents) section as shown in Figure 5

Training examples of wit.ai Classifier are given in Understanding as shown in Figure 6. Each sentence is annotated with an intent and the entities linked to that intent.

III. Dataset

We have generated our own French dataset in a recruitment interview context.

A. Data Generation

We created a list of 22 intents and generated a total of 2,547 examples:

- in varying words and syntaxes.

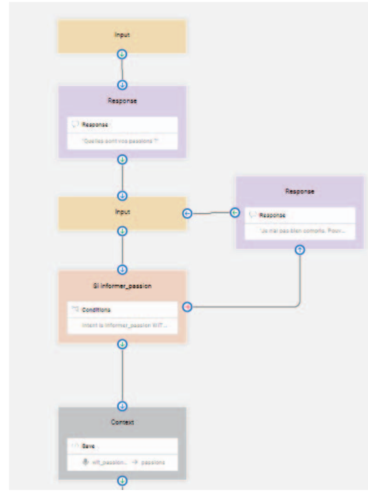

Fig. 4. Wit.ai Composer example.

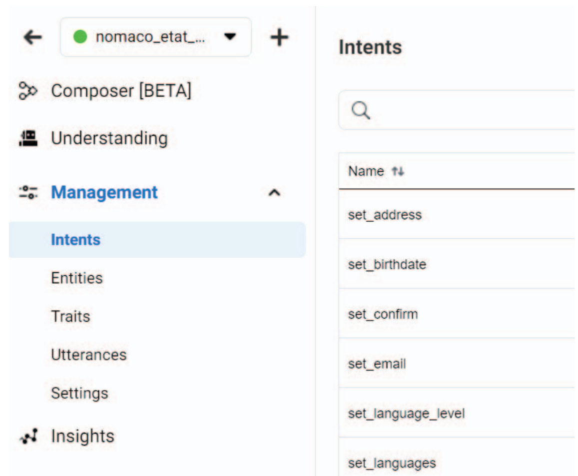

Fig. 5. Intents management.

- in varying completeness of information. Example: to not give the postal code for the "postal address" intent (informer_adressePostale in french). The objective is that the chatbot detects intents in the case of incomplete information.

The Figure 7 below shows the intents to be detected by the recruitment chatbot.

B. Data Annotation

The annotation is carried out in an Excel table by selecting an intent from the defined intent list presented in Figure 7.

Figure 8 shows a sample of generated and annotated data.

IV. Experiments

The aim of the experiments is to compare the three classifiers: DIETClassifier (Rasa), wit.ai Classifier and CamemBERT in the case of a French recruitment chatbot. Camem-

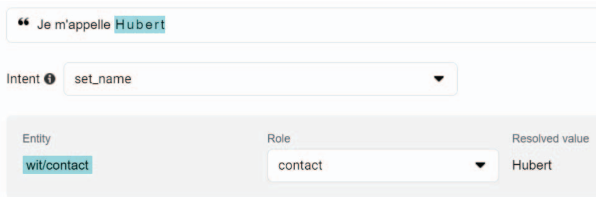

Fig. 6. Intent annotation example in wit.ai.

Intentions	
confirmer_entretien	informer_dilplome
refuser_entretien	informer_formation
poser_question	informer_certification
se_présenter	nier_diplome
informer_dateNaiss	nier_formation
informer_adressePostale	nier_certification
informer_tel	informer_expérience
informer_mail	nier_expérience
informer_permis	informer_hard_skills
refuser_répondre	informer_soft_skills
informer_langue	informer_loisir

Fig. 7. Intents list.

BERT [16] is a multi-layer bidirectional Transformer encoder. It is based on RoBERTa [17], which is based on BERT [18].

Dataset is divided into train and test set as shown in Figure 9.

We evaluated the models with the accuracy, recall and f1-score metrics and the confusion matrix.

A. DIETClassifier (Rasa)

We trained DIETClassifier from 100 to 500 epochs with a step of 50. The best accuracy obtained is 0.8668 with 200 epochs as shown in Figure 10.

The precision, recall and f1-score metrics and the confusion matrix are given in Figure 11 and Figure 12 respectively.

B. Wit.ai Classifier

wit.ai is not an open-source platform. The training parameters such as the number of epochs can not be modified. So, we trained the wit.ai Classifier with the basic parameter configuration. The precision, recall and f1-score metrics and the confusion matrix are given in Figure 13 and Figure 14 respectively.

The results show that the average accuracy of the wit.ai Classifier is lower than the accuracy of DIETClassifier by 2% with a much shorter execution time.

C. CamemBERT

We trained CamemBERT from 5 to 40 epochs with a step of 5. The average accuracy for each model is given in Figure 15. The best accuracy obtained is 0.93 with 5 epochs.

The precision, recall and f1-score metrics and the confusion matrix are given in Figure 16 and Figure 17 respectively.

Experiments have shown that DIETClassifier needs to be trained up to 200 epochs to reach its maximum performance.

	Texte	Intentions
T1	je suis léa martinez	se_présenter
T2	moi c'est Antoine leclerc	se_présenter
T3	Je m'appelle Asmaa Ben Moussa	se_présenter
T4	C'est lucas de la porte	se_présenter
T5	Micheal bois	se_présenter
T6	c'est Madame Emanuelle Petit	se_présenter
T7	Mon nom est Frank et mon prénom est Philippe	se_présenter
T8	je me présente, je suis Dr Houssem Najar	se_présenter
T9	Je suis Mr Clément léo	se_présenter
T10	Moi je m'apelle catrine levier	se_présenter

Fig. 8. A sample of our dataset.

train set	test set	total
2104	443	2547
83%	17%	100%

Fig. 9. Train and test set.

DIETClassifier									
Nbr epoch	100	150	200	250	300	350	400	450	500
Accuracy	0,8397	0,8532	**0,8668**	0,8623	0,8623	0,8645	0,8487	0,8487	0,8487

Fig. 10. DIETClassifier accuracy per epoch.

	precision	recall	f1-score	Support
confirmer_entretien	0,86	0,86	0,86	21
informer_adressePostale	0,95	1,00	0,98	21
informer_certification	0,90	0,95	0,92	19
informer_dateNais	1,00	0,89	0,94	18
informer_diplome	0,81	1,00	0,89	22
informer_experience	1,00	0,45	0,62	20
informer_formation	0,81	0,95	0,88	22
informer_hard_skill	0,61	0,85	0,71	20
informer_langue	0,91	1,00	0,95	20
informer_loisir	0,83	1,00	0,91	20
informer_mail	1,00	1,00	1,00	21
informer_permis	0,95	0,95	0,95	21
informer_soft_skill	0,86	0,60	0,71	20
informer_tel	1,00	0,93	0,97	15
nier_certification	1,00	0,90	0,95	20
nier_diplome	1,00	0,74	0,85	23
nier_experience	0,74	0,94	0,83	18
nier_formation	1,00	0,75	0,86	20
poser_question	0,83	1,00	0,91	20
refuser_entretien	0,72	0,86	0,78	21
refuser_repondre	0,81	0,65	0,72	20
se_présenter	0,90	0,90	0,90	21
accuracy			0,86	443

Fig. 11. DIETClassifier: precision, recall and f1-score metrics.

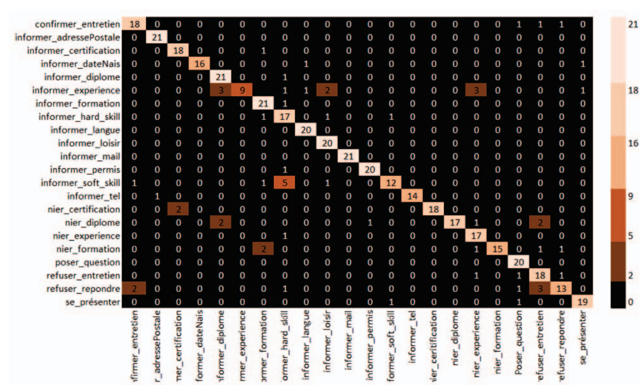

Fig. 12. DIETClassifier: confusion matrix.

	precision	recall	f1-score	Support
confirmer_entretien	0,95	0,90	0,93	21
informer_adressePostale	1,00	1,00	1,00	21
informer_certification	0,94	0,79	0,86	19
informer_dateNais	1,00	1,00	1,00	18
informer_diplome	0,92	1,00	0,96	22
informer_experience	1,00	0,45	0,62	20
informer_formation	0,75	0,95	0,84	22
informer_hard_skill	0,72	0,90	0,80	20
informer_langue	1,00	1,00	1,00	20
informer_loisir	0,95	0,90	0,92	20
informer_mail	1,00	1,00	1,00	21
informer_permis	0,95	1,00	0,98	21
informer_soft_skill	0,68	0,75	0,71	20
informer_tel	1,00	1,00	1,00	15
nier_certification	0,91	1,00	0,95	20
nier_diplome	1,00	0,74	0,85	23
nier_experience	0,80	0,89	0,84	18
nier_formation	0,89	0,85	0,87	20
poser_question	0,80	1,00	0,89	20
refuser_entretien	0,75	0,71	0,73	21
refuser_repondre	0,86	0,60	0,71	20
se_présenter	0,81	1,00	0,89	21
accuracy			0,88	443

Fig. 13. wit.ai: precision, recall and f1-score metrics.

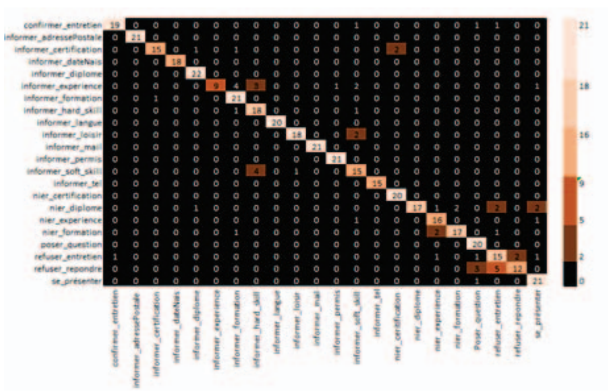

Fig. 14. wit.ai: confusion matrix.

CamemBERT								
Nbr epoch	5	10	15	20	25	30	35	40
Accuracy	**0,9300**	0,9142	0,9209	0,9074	0,9209	0,9187	0,9119	0,9119

Fig. 15. CamemBERT accuracy per epoch.

	precision	recall	f1-score	support
confirmer_entretien	0.81	1.00	0.89	21
informer_adressePostale	1.00	1.00	1.00	21
informer_certification	0.90	0.95	0.92	19
informer_dateNais	1.00	1.00	1.00	18
informer_diplome	0.88	1.00	0.94	22
informer_experience	1.00	0.95	0.97	20
informer_formation	0.91	0.95	0.93	22
informer_hard_skill	0.78	0.90	0.84	20
informer_langue	1.00	1.00	1.00	20
informer_loisir	0.95	1.00	0.98	20
informer_mail	1.00	1.00	1.00	21
informer_permis	1.00	1.00	1.00	21
informer_soft_skill	0.94	0.75	0.83	20
informer_tel	1.00	1.00	1.00	15
nier_certification	1.00	0.95	0.97	20
nier_diplome	0.86	0.83	0.84	23
nier_experience	0.94	0.83	0.88	18
nier_formation	1.00	0.85	0.92	20
poser_question	0.95	1.00	0.98	20
refuser_entretien	1.00	0.67	0.80	21
refuser_repondre	0.79	0.95	0.86	20
se_présenter	1.00	1.00	1.00	21
accuracy			0.93	443

Fig. 16. CamemBERT: precision, recall and f1-score metrics.

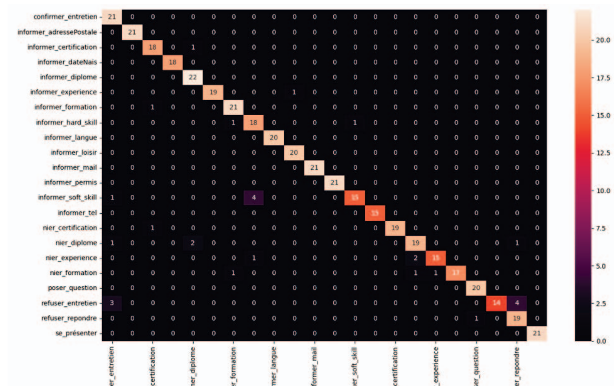

Fig. 17. CamemBERT: confusion matrix.

However, its performance remains below that of CamemBERT and wit.ai Classifier. CamemBERT ensures very good intent classification with an average accuracy of 93%, compared with 88% for wit.ai and 86% for DIETClassifier. For our French recruitment chatbot, CamemBERT is the best Classifier.

Moreover, experiments show that there are 2 ambiguities in the 3 models:

- between intents *inform soft skill* and *inform hard skill*: this is due to the fact that both intentions express skills, and the context of the sentence can easily be identical. Following example shows the confusion that can exist between the two intents.
  - I can make important decisions (soft skill intent)
  - I can solve management software problems (hard skill intent)
- between *decline interview* and *refuse to answer*: in both intents, we have an expression of refusal, which explains the confusion as shown in the following example.
  - I do not wish to take the interview (decline interview intent)
  - I do not wish to answer this question (refuse to answer intent)

CamemBERT's performance is explained by the fact that Camembert was specifically designed to capture the linguistic features and nuances of the French language. The models in Rasa and wit.ai are designed for natural language processing in the context of chatbots and virtual assistants.

Moreover, Camembert is pre-trained on a wide variety of French texts, while the models in the frameworks are pre-trained on simulated dialogues, diversified texts and real conversations.

V. Conclusion and Perspective

In this article, we present a comparative study of three classifier models: DIETClassifer, wit.ai and CamemBERT, for French language in intent classification task in the case of a chatbot interview job. To the best of our knowledge, there is no comparison of these models in the literature for French language. We generated a dataset in our context to train and test the models. The best classifier for our chatbot is CamemBERT which obtained an average accuracy of 93%. As a perspective, we plane to compare these results with the latest GPT models in intent classification and also entity extraction during a job interview with the chatbot.

REFERENCES

[1] B. R. Ranoliya, N. Raghuwanshi, and S. Singh, "Chatbot for university related faqs," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2017, pp. 1525–1530.

[2] H. B. Essel, D. Vlachopoulos, A. Tachie-Menson, E. E. Johnson, and P. K. Baah, "The impact of a virtual teaching assistant (chatbot) on students' learning in ghanaian higher education," *International Journal of Educational Technology in Higher Education*, vol. 19, no. 1, pp. 1–19, 2022.

[3] P. Amiri and E. Karahanna, "Chatbot use cases in the covid-19 public health response," *Journal of the American Medical Informatics Association*, vol. 29, no. 5, pp. 1000–1010, 2022.

[4] M. R. King, "The future of ai in medicine: a perspective from a chatbot," *Annals of Biomedical Engineering*, vol. 51, no. 2, pp. 291–295, 2023.

[5] Z. Chen, "Collaboration among recruiters and artificial intelligence: removing human prejudices in employment," *Cognition, Technology & Work*, vol. 25, no. 1, pp. 135–149, 2023.

[6] S. Koivunen, S. Ala-Luopa, T. Olsson, and A. Haapakorpi, "The march of chatbots into recruitment: recruiters' experiences, expectations, and design opportunities," *Computer Supported Cooperative Work (CSCW)*, vol. 31, no. 3, pp. 487–516, 2022.

[7] C. de Ruijt and S. Bhulai, "Job recommender systems: A review," *arXiv preprint arXiv:2111.13576*, 2021.

[8] N. Boudjani, V. Colas, C. Joubert, and D. B. Amor, "Ai chatbot for job interview," in *2023 46th MIPRO ICT and Electronics Convention (MIPRO)*. IEEE, 2023, pp. 1155–1160.

[9] "Chatgpt," https://openai.com/chatgpt, accessed: 2023-08-02.

[10] "Rasa," https://rasa.com/, accessed: 2023-08-02.

[11] "Wit.ai," https://wit.ai/, accessed: 2023-08-02.

[12] R. K. Sharma and M. Joshi, "An analytical study and review of open source chatbot framework, rasa," *Int. J. Eng. Res*, vol. 9, no. 06, pp. 1011–1014, 2020.

[13] A. Jiao, "An intelligent chatbot system based on entity extraction using rasa nlu and neural network," in *Journal of physics: conference series*, vol. 1487, no. 1. IOP Publishing, 2020, p. 012014.

[14] A. A. Qaffas, "Improvement of chatbots semantics using wit. ai and word sequence kernel: Education chatbot as a case study," *International journal of modern education and computer science*, vol. 11, no. 3, p. 16, 2019.

[15] C. Blanc, A. Bailly, É. Francis, T. Guillotin, F. Jamal, B. Wakim, and P. Roy, "Flaubert vs. camembert: Understanding patient's answers by a french medical chatbot," *Artificial Intelligence in Medicine*, vol. 127, p. 102264, 2022.

[16] L. Martin, B. Muller, P. J. O. Suárez, Y. Dupont, L. Romary, É. V. de La Clergerie, D. Seddah, and B. Sagot, "Camembert: a tasty french language model," *arXiv preprint arXiv:1911.03894*, 2019.

[17] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[18] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.