

Pierre ALAIN

pierre.alain@irisa.fr

Fabien AUTREL

fabien.autrel@imt-atlantique.fr

Guillaume DOYEN

guillaume.doyen@imt-atlantique.fr

ANNÉE : 2024

Du langage naturel aux réseaux basés sur les intentions

Master Recherche - Équipe SOTERN

1 CONTEXTE

Les réseaux basés sur les intentions ont pour but de mettre l'accent sur les intentions d'un administrateur plutôt que les moyens de configuration d'un réseau. Les intentions peuvent s'exprimer sous forme de haut niveau, par exemple un administrateur pourrait vouloir exprimer : « Les utilisateurs du département Ressources Humaines doivent pouvoir s'échanger des données de manière sécurisée. ». Cette intention doit alors pouvoir être transcrite dans une configuration du réseau et des équipements.

Depuis une décennie, les avancées des travaux sur les modèles de langage, et en particulier les larges (en anglais *LLM (Large Language Model)*, typiquement chatGPT pour OpenAI, Llama pour Facebook), permettent d'exprimer sous une forme d'un plongement (*embedding* en anglais) l'information d'un texte écrit en langage naturel (par exemple une description haut niveau du besoin de l'administrateur). Le plongement représente alors une forme condensée de l'entrée donnée au réseau (le nombre de dimensions (*features*) de l'entrée, pouvant être choisi arbitrairement grand, est réduit au niveau de l'*embedding* (c'est l'apprentissage du réseau qui se charge d'établir les correspondances les plus utiles dans l'*embedding*) et partir de cette donnée rend la tâche de traduction vers autre chose plus simple à apprendre).

Il n'existe pas encore de modèle de langage permettant d'effectuer la traduction depuis le langage naturel vers une description de réseau ou de politique de sécurité. Cependant, si cela existait, des problématiques seraient encore non résolues, la tâche de traduction langage naturel \rightarrow politique de sécurité du réseau permet-elle de garantir une correspondance exacte ? Les exigences de sécurité sont-elles préservées au cours de la traduction ?

2 PROBLÉMATIQUE ET OBJECTIFS

Les réseaux basés sur les intentions et les problématiques associées sont décrits dans [1]. Un système de réseau basé sur les intentions va reposer sur plusieurs étapes :

1. la détection et la compréhension de l'intention
2. la traduction de l'intention
3. la détection et résolution éventuelle de conflit dans les intentions
4. l'activation de l'intention
5. le respect de l'intention dans le réseau

Le premier point, concernant la détection et la classification des intentions est un sujet encore ouvert, en particulier via l'utilisation de réseaux de neurones [2]. Les modèles de langage larges utilisent plusieurs millions ou milliards de paramètres et ont été appris sur le langage naturel [3].

Le second point peut être assimilé à une tâche de traduction [4]. Dans le cadre des réseaux basés sur les intentions, il existe plusieurs langages de description de réseau basé sur les intentions, parmi ceux-ci on peut citer le langage NILE (Network Intent Language) [5] utilisant des réseaux de neurones sequence-to-sequence pour réaliser la traduction du langage de haut niveau vers le langage NILE et proposent une plateforme complète dans [6] (la présentation vidéo est disponible¹).

L'**objectif du stage** porte sur les deux premières tâches d'identification des intentions et de traduction depuis un langage de haut niveau vers un langage permettant d'en dériver des politiques de sécurité ou de configuration d'équipements.

Concernant le troisième point [7] apporte des pistes sur la résolution d'un ensemble de contraintes pouvant être contradictoires. Le corpus mis à disposition avec [6] donne des exemples de règles en conflits.

¹<https://www.usenix.org/conference/atc21/presentation/jacobs>

3 TÂCHES À RÉALISER

Le but du stage porte sur l'évaluation des techniques développées dans le cadre de [6] et l'utilisation des techniques neuronales ailleurs qu'uniquement dans la reconnaissance des entités nommées. Est-il possible de réaliser une traduction depuis le langage naturel vers l'intention dans le langage NILE ?

Les corpus d'apprentissage de la plateforme LUMI sont disponibles à <https://github.com/lumichatbot/webhook/tree/master/res/dataset>. Les tâches à réaliser seront de partir du corpus mis à disposition pour [6], et d'utiliser un réseau de neurones pour réaliser une tâche de traduction depuis le langage naturel vers le langage de description NILE. Le corpus d'apprentissage disponible dans le dépôt de LUMI peut se révéler trop petit pour les besoins des modèles de traduction, aussi peut-être nécessaire d'utiliser des techniques d'augmentations [8].

Afin de simplifier la partie de développement, avec la grande majorité des prototypes de la communauté des réseaux de neurones disponibles en python, les développements seront à réaliser dans ce langage.

BIBLIOGRAPHIE

- [1] A. Leivadreas, and M. Falkner, "A survey on intent-based networking," *IEEE Commun. Surv. Tut.*, vol. 25, no. 1, pp. 625–655, 2023, doi: 10.1109/COMST.2022.3215919. [Online]. Available: <https://doi.org/10.1109/COMST.2022.3215919>
- [2] D. Comi, D. Christofidellis, P. F. Piazza, and M. Manica, "Z-bert-a: a zero-shot pipeline for unknown intent detection," 2022.
- [3] H. Touvron, T. Lavril, et al., "Llama: open and efficient foundation language models," 2023.
- [4] W. Zhu, H. Liu, et al., "Multilingual machine translation with large language models: empirical results and analysis," 2023.
- [5] A. Jacobs, R. Pfitscher, R. Ferreira, and L. Granville, "Refining network intents for self-driving networks," 2018, pp. 15–21, doi: 10.1145/3229584.3229590.
- [6] A. S. Jacobs, R. J. Pfitscher, et al., "Hey, lumi! using natural language for Intent-Based network management," in *2021 USENIX Annu. Tech. Conf. (USENIX ATC 21)*, Jul. 2021, pp. 625–639. [Online]. Available: <https://www.usenix.org/conference/atc21/presentation/jacobs>
- [7] D. D. A. Nguyen, F. Autrel, A. Bouabdallah, and G. Doyen, "A robust approach for the detection and prevention of conflicts in i2nsf security policies," in *NOMS 2023-2023 IEEE/IFIP Netw. Operations Manage. Symp.*, vol. 0, 2023, pp. 1–7, doi: 10.1109/NOMS56928.2023.10154304.
- [8] T. Dopierre, C. Gravier, and W. Logerais, "Protaugment: unsupervised diverse short-texts paraphrasing for intent detection meta-learning," *Arxiv*, 2021.