Projet de Recherche ANRT
Few vs Zero shot learning for Traffic Monitoring/Management

PhD Candidate: Chao Wang

Academic Supervisor: Prof. Pietro Michiardi, Ph.D.

Industrial Supervisor: Dr. Lixuan Yang, Ph.D.

1 Introduction

Traffic monitoring is a strategic activity for networks operation & management (O&M). A variety of methodologies and technologies have been developed to shed light on both the network routers and health of Internet services. Many of such tools are based on the adoption of "classifiers", i.e., entities that are able to label traffic properties according to predefined categories. In particular, *traffic classification* methods label network flows with the Internet application generating them, thus enabling both in depth view of the traffic composition (e.g., study which are the popular services for end users) enabling an advanced management of their performance (e.g., prioritize gaming applications over mail services). Thus, good classifiers are vital to provide superior network O&M.

The recent bloom of Artificial Intelligence methods in computer vision and natural language processing is paving the way to the next generation of traffic monitoring. However, there are still many open challenges before achieving the vision of *self driving networks* [1]. In particular, network traffic data is highly dynamic and dependent on the network environment generating it – traffic monitoring is strongly *environment dependent*. For instance, while Deep Learning (DL) enables the creation of models successfully handling hundreds of network applications [2], such models are only well fit for the environment that they have been trained on. This implies that solutions developed by one operator cannot be easily ported to another operator, as the set of applications may differ, or behave differently in the different environment. Likewise, within

a single network operator, solutions built for residential access networks do not necessarily apply to enterprise or campus networks. As such, by relying only on current state of the art methodologies in the networking community, the data collected would lead to models that can only apply to the networks for which they have been trained on.

This problem is further exacerbated by the rapid evolution of Internet traffic and services: as pointed out in[3], models life-cycle management require to continuously adapt models to the operational environment, which requires to iterate between Ops and Dev stages. Since traffic classifiers are based on supervised techniques, such models life-cycle requires in principle a continuous stream of labeled data. Unfortunately collecting such labeled dataset requires huge efforts and the process can be entangled with data privacy/confidentiality issues. Thus, efficient traffic monitoring not only requires effective classifiers, highly adaptive to the operational environment they are expected to be deployed, but they should also operate with very *little labeled data*.

This calls for an *higher level of abstraction* to be learned from the data – rather than starting new models training at each new environment or environmental change (i.e., to learn how to identify specific patterns related to the dataset at hand), the desirable methodologies should be able to learn (i) more general properties, with as little data as possible, that (ii) can be quickly adapted to different environments.

2 Thesis foundation and methodologies

This PhD thesis project centers around the study of novel techniques to obtain classifiers for traffic monitoring that are (*i*) highly portable across network environments and (*ii*) require as little as possible training data and/or labels. We believe the combination of these two aspects to be at the foundation of AI model life-cycle management.

This calls for research into *Zero-shot learning* (ZSL) and *Few shot learning* (FSL), that can enable a higher level of abstraction to describe model classes – i.e., to achieve *better representation* of the data.

In particular, FSL (or ZSL) is a recent research area which achieved promising results in the computer vision community [4], and is based on the idea of creating a classifier that can properly detect classes, even when trained with few (or none) examples of the new class. Whereas the FSL methodology enables learning to generalize from very few examples (generally few units/tens), the ZSL method is even more extreme (e.g., identify birds without being trained with labeled instances of birds). As in ZSL no labeled instances belonging to the unseen classes are available, *auxiliary information* is necessary. Each class is associated with semantic information, namely *class prototypes* formulated in different ways: engineered spaces (flow attributes, text-keywords to describe applications) and learned space (labeled flows, traffic representation via an embedding). While FSL and ZSL are the basic building blocks of the Thesis, we point out that several other natural language processing techniques can be

leveraged to achieve a better semantic representation of prototypes, e.g, word embedding [5], attention models [6] and Transformers [7] or Graph Neural Networks [8, 9], which exploit relationship between attributes.

We note that FSL and ZSL have only been studied in the context of image processing, while to the best of our knowledge, those have not been applied to network data yet. We also underline that, differently from image data, network data exhibit patterns which are tightly related to the nature of network protocols. Yet, such native property has not yet been fully explored in the context of Deep Learning techniques, while we believe they can have a important role when defining better data representations for network data.

2.1 Workplan

Huawei has a private commercial-grade dataset [2] which are collected from real networks including a very large set of applications. This will considerable speedup the research work and will be made available to the community (if possible). This dataset will enable to train a model to learn a rich latent representation including all possible (known) traffic attributes. This representation is critical for learning new classes from higher level attributes. In addition, for scientific reproducibility reasons, the workplan will unfold on experiments that use publicly available network traces, and eventually other kinds of data that share similar traits to network traffic.

The three years PhD program can be organised as following:

- T00-T04 Study the state of the art on FSL/ZSL (including representation learning).
- T04-T12: Few shot learning (on existing datasets).
- T12-T16: Augmenting datasets, and preliminary representation learning solution.
- T16-T24: Zero shot learning (on augmented dataset).
- T24-T32: Comparison of FSL vs ZSL techniques.
- T32-T36: Thesis drafting and manuscript preparation.

Milestones associated to the different phases include:

- (M1, T04): a state of the art survey (that serve as a basis for the research and one chapter of the thesis, but could be worth of a standalone submission)
- (M2, T12): technical report on FSL on existing dataset (submission)
- (M3, T16): technical report on effective representation learning techniques (dataset chapter of the thesis)
- (M4, T24): technical report on ZSL on augmented dataset (submission)

- (M5, T32): technical report comparing pros and cons of FSL vs ZSL on a head-to-head basis (submission).

For M4, a patent submission is also possible in alternative or addition to a scientific publication, depending on the obtained results.

3 Résumé en français

La surveillance du trafic est une activité stratégique pour l'opération & la gestion des réseaux (O&M). La majorité de ces outils sont basé sur des modèles classification pour étiqueter les trafics selon des catégories prédéfinies. Les systèmes de *classification du trafic* permettant ainsi à la fois comprendre de la composition du trafic (par exemple, l'étude des services populaires pour les utilisateurs). De plus, ces modèles permettent une gestion avancée de leurs performances (par exemple, donner la priorité aux applications de jeu vidéo par rapport aux services de messagerie). Ainsi, une classificateur robusts est essentiel pour fournir un réseau de qualité supérieur O&M.

Ce projet de thèse s'articule autour de l'étude de nouvelles techniques permettant d'obtenir des classificateurs pour la surveillance du trafic qui sont (i) hautement portables dans les environnements réseau et (ii) nécessitent le moins possible de données d'entraînement et/ou d'étiquettes. La combinaison de ces deux aspects est la base de la gestion du cycle de vie pour les modèles d'intelligence artificielle (IA).

Cette objective nécessites des recherches sur *Zero-shot learning* (ZSL) et *Few shot learning* (FSL), qui peuvent obtenir un niveau d'abstraction plus élevé pour décrire les classes de modèles - c'est-à-dire pour obtenir *meilleure représentation* des données.

En particulier, FSL (ou ZSL) est un domaine de recherche récent qui a obtenu des résultats prometteurs dans la communauté de la vision par ordinateur [4], et qui est basé sur la création d'un classificateur capable de détecter correctement les classes, même lorsqu'elles sont entraînées avec peu (ou aucune) exemples de la nouvelle classe. Alors que la méthodologie FSL permet d'apprendre à généraliser à partir de très peu d'exemples (généralement peu d'unités/dizaines), la méthode ZSL est encore plus extrême (par exemple, identifier les oiseaux sans être entraînés avec des instances étiquetées d'oiseaux). Pour ZSL, aucune instance étiquetée de nouvelles classes n'est disponible, *informations auxiliaires* est nécessaire. Chaque classe est associée à une information sémantique, à savoir *class prototypes* formulés de différentes manières : espaces ingénieries (attributs de flux, mots-clés-textes pour décrire des applications) et espace appris (représentation du trafic flow2vec).

Huawei dispose d'un ensemble de données privés [2] qui sont collectées à partir de réseaux réels comprenant un très grand nombre d'applications. Cela accélérera considérablement le travail de recherche et sera mis à la disposition de la communauté (si possible). Cet ensemble de données permettra d'entraîner un modèle pour apprendre une représentation riche comprenant tous les attributs de trafic possibles (connus). Cette représentation est essentielle pour apprendre

de nouvelles classes à partir d'attributs de niveau supérieur. En outre, pour des raisons de reproductibilité scientifique, le plan de travail se déroulera sur des expériences utilisant des traces de réseau accessibles au public, et éventuellement d'autres types de données partageant des caractéristiques similaires au trafic réseau.

References

[1] N. Feamster and J. Rexford, "Why (and how) networks should run themselves," *arXiv:1710.11583*, 2017.

[2] L. Yang, A. Finamore, J. Feng, and D. Rossi, "Deep learning for encrypted zero-day traffic classification," *arXiv 2104.03182*, 2021.

[3] L. Yang and D. Rossi, "Quality monitoring and assessment of deployed deep learning models for network aiops," *arXiv 2104.03182*, 2021.

[4] W. Wang, V. Zheng, H. Yu, and C. Miao, "A survey of zero-shot learning: Settings, methods, and applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, pp. 1–37, 01 2019.

[5] J. Pennington, R. Socher, and C. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. [Online]. Available: https://aclanthology.org/D14-1162

[6] Y. Liu, J. Guo, D. Cai, and X. He, "Attribute attention for semantic disambiguation in zero-shot learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *CoRR*, vol. abs/1706.03762, 2017. [Online]. Available: http://arxiv.org/abs/1706.03762

[8] Y. Geng, J. Chen, Z. Chen, Z. Ye, Z. Yuan, Y. Jia, and H. Chen, "Generative adversarial zero-shot learning via knowledge graphs," *CoRR*, vol. abs/2004.03109, 2020. [Online]. Available: https://arxiv.org/abs/2004.03109

[9] Y. Geng, J. Chen, Z. Chen, J. Z. Pan, Z. Yuan, and H. Chen, "K-ZSL: resources for knowledge-driven zero-shot learning," *CoRR*, vol. abs/2106.15047, 2021. [Online]. Available: https://arxiv.org/abs/2106.15047