Session 5: Session 5: ML for networking – can “transformers” transform networks?

Transcribed by: M. Taimoor Tariq, Ammar Tahir (UIUC)

A New Hope for Network Model Generalization

Alexander Dietmüller*, Siddhant Ray, Romain Jacob, Laurent Vanbever (ETH Zürich)*

Transfer learning allows generalization of ML models where an ML model trained on a large dataset can be repurposed with a small amount of training. This paper applies transformers to networks to enable an ML model that can generalize across diverse packet traces.

Questions:

1. **Intuition for transformers is that it relies on grammar, or some sort of structure. What structure do we have for networks? What is the notion of grammars do we have that makes it usable in networks?**
   There are fundamental patterns in networks. Congestion builds up slowly, wireless networks queue up packets in certain patterns, these underlying patterns are learnt by transformers who can pick up on these transformers.

Towards a systematic multi-modal representation learning for network data

Zied Ben Houidi*, Raphael Azorin, Massimo Gallo, Alessandro Finamore, Dario Rossi (Huawei Technologies)*

Network data is difficult to model, has various modalities with no unified format. We need to find a unifying format which covers all the different modalities.

Questions:

1. **Multi modal learning is that you are learning with 2 different types of data and how much variation for the data can you have for each modality and how do you make sure it's enough?**
   Once we define the model for our model, it becomes a classic ML problem and we can adopt classic ML techniques to compose a model. We use access point embedding, the users do not move randomly and this information we encode in the AP embeddings and the same principle applies to many other use cases as well.
2. **How robust is your approach in terms of (multiple) missing modalities?**
   For missing modalities , Good question but this is something for future work. Maybe we can generate one modality from another modality, e.g. generate image from text or vice versa.

Rethinking Data-driven Networking with Foundation Models: Challenges and Opportunities

***Franck Le**, Mudhakar Srivatsa, Raghu Ganti (IBM); Vyas Sekar (Carnegie Mellon University)*

Foundational models with contextual embeddings are able to assign unique meanings even to the same words in the same sentence.
Similar to NLP models, hope is that the foundational models can learn the relationships between different networking components, e.g. king similar to queen, TCP similar to UDP (both are transport protocols)
So the simple question is if FM can help networking.

Question:

1. **Analogies with the English language can only go so far? Sentiment analysis to network components sounds somewhat impractical.**
   If we apply say an RNN, compared to the Foundational model can boost F1 score significantly. It does seem promising. In sentiments there's a lot of uncertainty, but in networking it can be much more deterministic.

End of Session Discussion

1. **Network traces are very long compared to textual data, how is that a challenge?**
   Transformers struggle with long data, in most cases we need more information about the recent packet compared to older packet, so older packets can be aggregated. So aggregation can turn long sequences into shorter manageable sequences.
2. **Pre Training on these models. Whether it's about data for pretraining on or objective for pretraining. How can we generalize the models, e.g. from universities network to the general networks.**
   We can narrow down what we really want to learn about. Because in general networks you have to learn about user behavior as well as traffic behavior. We take the user behaviors, and see how the network responds to that. This is not a fundamental limitation but for the first prototype we felt this was a wise choice to get reasonable results.
3. **Network traffic changes much more quickly than English? What are the safety guarantees?**
   -Fundamental changes do not come up that often. We need to consider updating models but it may be long enough that they are useful.
   - Continuously add new embeddings to update the model to keep up.

4. **Assumptions about the data being used, and if we dump all the data to model there will be a low signal to noise ratio. So how to curate data to produce practical results. Aren't you assuming that the data that you have is useful and reliable?**
   The primary goal is to make this system work without human intervention by standardizing the way data is represented.
   We are also banking on the fact that transformers can pick out the network dynamics in different contexts. WE think that the network has the capability and figure out and share the dynamics.
   It would be interesting to see as you degrade the quality of data, how the performance of the model deteriorates.
5. **What do you do about outliers? Oftentimes they are of interest but learning-based solutions can ignore 1% outliers.**
   Simulation with large variance. However, when the model is biased you can only learn that much so this is a problem. We need to ensure diversity in training data so we do not pick just majority patterns.
6. **Forgetfulness in the model?**
   Not a specific problem but general to all machine learning . Important to have diversity in the training data to ensure as much diversity of features as possible so we dont forget things when we don't have much data left to learn from.