PhD Topic: Representation Learning for Privacy-Preserving Speech Recognition

Aurélien Bellet, Marc Tommasi, Emmanuel Vincent

February 7, 2018

Teams and contact

- Magnet team, Inria/CRIStAL (Lille, France): http://team.inria.fr/magnet
  Aurélien Bellet (aurelien.bellet@inria.fr), Marc Tommasi (marc.tommasi@inria.fr)
- Multispeech team, Inria/Loria (Nancy, France): http://team.inria.fr/multispeech
  Emmanuel Vincent (emmanuel.vincent@inria.fr)

Keywords

Machine Learning, Speech Recognition, Privacy, Representation Learning, Deep Learning.

Context

With the advent of the big data era, great progress has been made in automatic speech recognition over the last decade [8, 7]. This is due to the maturity of machine learning techniques (e.g., advanced forms of deep learning such as very deep convolutional neural networks [2] or attention-based encoder-decoder architectures [3]) but also to the availability of very large datasets and the increase in computational power. Speech recognition is now embedded in many applications running on various devices (computers, smartphones, smart home systems, GPS...) in the form of virtual assistants proposed by a few big players.[1] The users of such applications implicitly participate in the data collection process as their speech signals are recorded, sent and stored on central servers (huge clusters of computers owned by the above leaders of the digital economy) where they are used to train and update deep neural network models.

This centralization induces serious privacy concerns: speech data always contains private/sensitive information in the user's voice and sometimes in the spoken message itself. In the event of a security breach, this could be malevolently used by attackers to gather information about the user's lifestyle and health status, or to build a synthesized voice that impersonates him/her. So far, the design of privacy-preserving systems has not been much investigated in speech processing [6] and the proposed techniques are not applicable to deep learning.

Objectives

The goal of this PhD is to propose methods for privacy-preserving speech recognition: the device of each user does not share its raw speech data and runs some private computations locally, while some cross-user computations may be done by communicating through a server (without exposing sensitive information). More specifically, we will rely on representation learning to separate the features of the user data that can expose private information from generic ones useful for the task

[1] For instance: Apple's Siri, Google Now, Microsoft's Cortana, and Amazon's Alexa.

of interest. Several directions will be explored, among which adversarial deep learning methods such as adversarial autoencoders [5] as well as connections to multi-task learning [1] and to the emerging field of fairness in machine learning [9]. We will establish formal privacy guarantees such as differential privacy [4], and the proposed methods will be evaluated on standard speech recognition benchmark datasets such as Librispeech.[2]

References

[1] A. Argyriou, T. Evgeniou, and M. Pontil. Convex multi-task feature learning. *Machine Learning*, 73(3):243–272, 2008.

[2] M. Bi, Y. Qian, and K. Yu. Very deep convolutional neural networks for LVCSR. In *Proceedings of Interspeech*, pages 3259–3263, 2015.

[3] C.-C. Chiu, T. N. Sainath, Y. Wu, R. Prabhavalkar, P. Nguyen, Z. Chen, A. Kannan, R. J. Weiss, K. Rao, E. Gonina, N. Jaitly, B. Li, J. Chorowski, and M. Bacchiani. State-of-the-art speech recognition with sequence-to-sequence models. Technical report, arXiv:1712.01769, 2017.

[4] C. Dwork. Differential privacy: A survey of results. In *Proceedings of the International Conference on Theory and Applications of Models of Computation (TAMC)*, 2008.

[5] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey. Adversarial autoencoders. Technical report, arXiv:1511.05644, 2015.

[6] M. A. Pathak, B. Raj, S. D. Rane, and P. Smaragdis. Privacy-preserving speech processing: cryptographic and string-matching frameworks show promise. *IEEE Signal Processing Magazine*, 30(2):62–74, 2013.

[7] G. Saon, G. Kurata, T. Sercu, K. Audhkhasi, S. Thomas, D. Dimitriadis, X. Cui, B. Ramabhadran, M. Picheny, L.-L. Lim, B. Roomi, and P. Hall. English conversational telephone speech recognition by humans and machines. Technical report, arXiv:1703.02136, 2017.

[8] W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu, and G. Zweig. Achieving human parity in conversational speech recognition. Technical report, arXiv:1610.05256, 2017.

[9] R. S. Zemel, Y. Wu, K. Swersky, T. Pitassi, and C. Dwork. Learning fair representations. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2013.

[2] http://www.openslr.org/12/