

Proyecto de Tesis de Maestría (Miguel Bernal)

Título: Nuevas técnicas de traducción voz a voz basadas en inteligencia artificial.

Problema a resolver: La aplicación de la inteligencia artificial en áreas de ingeniería eléctrica exige una formación multidisciplinaria con una fuerte componente en lenguajes de programación y una distribución de los distintos módulos de trabajo según la especialización requerida [1]. La presente propuesta está dirigida a estudiantes de ingeniería en computación, informática o software, por concentrarse en el módulo correspondiente a la *traducción voz a voz de lenguaje natural*, como interfaz para sistemas robóticos [2]. La inteligencia artificial consiste en una serie de paradigmas de programación que imitan los sistemas biológicos para resolver problemas cuya formulación explícita es muy difícil o imposible [3]; en años recientes se ha dado en llamarle *machine learning* [4]. Consta de tres áreas principales: la *lógica difusa* que sistematiza el estudio de la incertidumbre [5], [6], las *redes neuronales* que representan conocimiento distribuido con base en algoritmos de aprendizaje [7], [8], y los algoritmos genéticos y evolutivos que resuelven problemas de competencia y optimización [9]. Desde 2006, la inteligencia artificial se ha visto renovada por novedosas técnicas del así llamado aprendizaje profundo (*deep learning*, [10]) que han aumentado notablemente su capacidad para resolver problemas de reconocimiento de imágenes, procesamiento de lenguaje [11], [12], minería de datos [13] y otros. Dichas técnicas han mejorado estructuras tradicionales como el *perceptrón* y algoritmos bien establecidos como el de *retropropagación* a través de *convolución*, *realimentación* y *recursión* [14], [15]. Dentro del problema de traducción voz a voz de lenguaje natural, las soluciones comúnmente empleadas pasan por las siguientes etapas [16], [17]: *análisis de características* (con métodos basados en redes neuronales o cadenas de Markov y herramientas tales como HTK o SPHINX, [18]), *detección de unidades* (como fonemas), *análisis sintáctico* (en caso de que se requiera validar las palabras) y *traducción* (respecto a una base de datos). Prácticamente no se han detectado enfoques basados solamente en el reconocimiento y aprendizaje fonético, *sin pasar por una base de datos*. Este trabajo se propone explorar esta posibilidad utilizando todas las herramientas de la inteligencia artificial (no estadística) a su alcance.

Objetivos: Desarrollar nuevos métodos de traducción voz a voz basados en inteligencia artificial.

- 1) Ilustrar las técnicas de traducción voz a voz existentes basadas en diccionarios e identificación de palabras.
- 2) Ilustrar la viabilidad de las técnicas de aprendizaje profundo para la traducción voz a voz con base de datos fonética.
- 3) Con base en los resultados anteriores, realizar modificaciones pertinentes a los algoritmos de aprendizaje para la mejora específica del enfoque voz a voz puramente fonético.

Productos académicos comprometidos: 1 artículo de conferencia internacional arbitrada publicado (Conference on Electrical Engineering, Computing Science and Automatic Control 2019 o Mexican Conference on Pattern Recognition 2020) y 1 artículo de revista indizada sometido, ambos antes del 31 de agosto de 2020.

Estancia del estudiante: En institución nacional por definir con duración de 1 mes.

REFERENCES

- [1] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited., 2016.
- [2] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.
- [3] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
- [4] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of machine learning research*, vol. 12, no. Aug, pp. 2493–2537, 2011.
- [5] L. Wang, *A Course in Fuzzy Systems and Control*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1997.
- [6] J. Kacprzyk and S. Zadrozny, "Computing with words is an implementable paradigm: fuzzy queries, linguistic data summaries, and natural-language generation," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 3, pp. 461–472, 2010.
- [7] S. Haykin, *Neural networks*. Prentice hall New York, 1994, vol. 2.
- [8] Y. Goldberg, "A primer on neural network models for natural language processing," *Journal of Artificial Intelligence Research*, vol. 57, pp. 345–420, 2016.
- [9] M. Gen and R. Cheng, *Genetic algorithms and engineering optimization*. John Wiley & Sons, 2000, vol. 7.
- [10] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [11] R. Collobert and J. Weston, "A unified architecture for natural language processing: Deep neural networks with multitask learning," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 160–167.
- [12] B. Bouchon-Meurio and G. Moysé, "Fuzzy linguistic summaries: Where are we, where can we go?" in *2012 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFER)*. IEEE, 2012, pp. 1–8.
- [13] M. Nafari and C. Weaver, "Query2question: translating visualization interaction into natural language," *IEEE transactions on visualization and computer graphics*, vol. 21, no. 6, pp. 756–769, 2015.
- [14] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, 2018.
- [15] D. Wang and J. Chen, "Supervised speech separation based on deep learning: An overview," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 10, pp. 1702–1726, 2018.
- [16] H. Krupakar, K. Rajvel, B. Bharathi, S. A. Deborah, and V. Krishnamurthy, "A survey of voice translation methodologies—acoustic dialect decoder," in *2016 International Conference on Information Communication and Embedded Systems (ICICES)*. IEEE, 2016, pp. 1–9.
- [17] J. Su, J. Zeng, D. Xiong, Y. Liu, M. Wang, and J. Xie, "A hierarchy-to-sequence attentional neural machine translation model," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 26, no. 3, pp. 623–632, 2018.
- [18] W. Walker, P. Lamere, P. Kwok, B. Raj, R. Singh, E. Gouvea, P. Wolf, and J. Woelfel, "Sphinx-4: A flexible open source framework for speech recognition," 2004.