Seminario sobre equidad en algoritmos

Dra.Viviana Erica Cotik (Profesora adjunta, Depto de Computación, FCEN-UBA)

Programa:
El aprendizaje automático se basa en que las computadoras aprendan a identificar patrones y tomar decisiones que imiten las de los humanos a partir de datos que se le presentan y con poca intervención humana. Actualmente el aprendizaje automático está integrado en nuestras vidas para la automatización de tareas de distinta índole y para la toma de decisiones.

Existen distintos tipos de sesgos, por raza y etnicidad, edad, sexo, clase, entre otros. Los algoritmos de aprendizaje automático aprenden a partir de los datos con los que se los entrena y pueden perpetuar los sesgos que existen en los mismos. De esta forma pueden otorgar oportunidades, recursos e información de manera injusta, proveyendo distinta calidad de servicio para distintas personas. También pueden reforzar estereotipos sociales

Es de suma importancia implementar algoritmos imparciales, usando técnicas de aprendizaje automático que consideren los sesgos y la equidad (el fairness). Para esto hay que considerar, entre otros, el muestreo de datos, el entrenamiento y la evaluación de los algoritmos, la composición de los equipos que construyen los sistemas (género, razas, disciplinas) y el procesamiento de datos. La implementación de algoritmos teniendo en cuenta la equidad es importante, entre otras áreas de estudio en el procesamiento del lenguaje natural, el procesamiento del habla y el procesamiento de imágenes.

En este seminario discutiremos el concepto de equidad (o imparcialidad) algorítmica y sesgos en sistemas de aprendizaje automático, sus posibles orígenes, la importancia de conseguir un aprendizaje automático justo y ético y algunas vías para avanzar en dicho sentido.

Se proporcionará a los estudiantes una lista de papers relevantes en la temática. Se pedirá presentación de los trabajos. Se propondrá preguntas para orientar las discusiones. También se elaborará un trabajo práctico.

Objetivos
En este seminario discutiremos el concepto de equidad (o imparcialidad) algorítmica y sesgos en sistemas de aprendizaje automático, sus posibles orígenes, la importancia de conseguir un aprendizaje automático justo y ético y algunas vías para avanzar en dicho sentido.

Temario:

- Introducción a la equidad en aprendizaje automático. Conceptos básicos de equidad (fairness) y sesgo en algoritmos. Desafíos éticos en aprendizaje automático.
- Sesgo en procesamiento del lenguaje natural y en procesamiento de imágenes y estrategias de abordaje.
- Seminarios sobre papers seleccionados que profundizan en temas específicos de fairness en algoritmos de aprendizaje automático, tanto en imágenes como en texto.
- Proyecto final: Desarrollo y evaluación de un modelo justo aplicado a un caso de estudio concreto.

Bibliografía:

- “Fair Enough: Standardizing Evaluation and Model Selection for Fairness Research in NLP. “Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, pp 297–312, 2023.
- “Inspecting Algorithms for Bias”. MIT Technology Review. 2017.
- “How We Analyzed the COMPAS Recidivism Algorithm.” J. Larson, S. Mattu, L. Kirchner, J. Angwin. ProPublica, 2016.
- “Fairness and Machine Learning". MIT Press. Solon Barocas, Moritz Hardt, Arvind Narayanan. 2023.
- “Discriminating Data”. Wendy Hui Kyong Chun. Charlas. https://www.youtube.com/watch?v=vnseu29xZHg, https://www.youtube.com/watch?v=AVMTGBecxfk
- "Causal Inference for Statistics, Social, and Biomedical Sciences", Imbens, Rubin. Cambridge University Press, 2015.
- “Fairness tutorial: CVPR2022 Fairness Tutorial”. Conference on Computer Vision and Pattern Recognition. Koyejo, Russakovsky.
- Tom Mitchell, "Machine Learning", McGraw Hill, 1997.
- Hastie, Tibshirani, Friedman, "The Elements of Statistical Learning", Springer, 2001.
- Bishop, "Pattern Recognition and Machine Learning", Springer, 2006.
- Jurafsky D, Martin JH. An introduction to natural language processing, computational linguistics, and speech recognition. Prentice Hall. 2024.
- “Neural Network Methods in Natural Language Processing (Synthesis Lectures on Human Language Technologies)”. Yoav Goldberg. Morgan & Claypool Publishers. 2017.
- “Introduction to Natural Language Processing”. Jacob Eisenstein. The MIT Press. 2019.
- “Foundations of Statistical Natural Language Processing”. Chris Manning , Hinrich Schütze. The MIT Press. 1999