

Proyectos

Hugo Jair Escalante

INAOE, CINVESTAV, México
ChaLearn, USA

hugo.jair@cs.cinvestav.mx

<https://ccc.inaoep.mx/~hugojaire/>

© Hugo Jair Escalante 2019



Objetivo

Desarrollar un proyecto de *investigación* en el cual se pongan en práctica conceptos y técnicas vistos durante el curso. El proyecto deberá responder a alguna pregunta de investigación, generar nuevo conocimiento.

Qué se espera de un proyecto de curso

- Asociado a un problema de relevancia local/global
- Se aplican conocimientos adquiridos en el curso para abordar el problema
- Involucra una etapa de investigación, se incorporan técnicas del SoTA (se hayan visto en clase o no)
- Se hace una contribución al conocimiento (de magnitud apropiada)

Qué involucra

- 1 Definición y delimitación del problema a abordar
- 2 Datos
- 3 Revisión del estado del arte (poner en contexto el trabajo / determinar su originalidad)
- 4 Hipótesis, determinar contribución,
- 5 Metodología propuesta
- 6 Diseño de experimentos (que preguntas de investigación se quieren resolver)
- 7 Hands on!
- 8 Análisis de resultados, explicar hallazgos
- 9 Conclusiones derivadas del trabajo

Friendly reminder....

Evaluación:

- Asistencia y participación - 10 %
- Exámenes y tareas - 30
- Proyecto final y avances - 60

Proyectos propuestos

- La siguiente es una lista de proyectos para el curso de Tópicos Selectos en Minería de Datos
- Los estudiantes seleccionarán un proyecto dentro del plazo establecido, o se les asignará a discreción
- Si tienen un proyecto/idea en mente, también es posible orientar el proyecto de clase en dicha dirección

Proyecto

Perfilado de usuarios en Twitter: A partir de los tuits de usuarios mexicanos, desarrollar métodos que permitan determinar la profesión del usuario y posiblemente otros aspectos (e.g., lugar de residencia, etc.) usando información textual y visual.

- 1 Desarrollo de métodos para el análisis de la información textual
- 2 Estudio e implementación de descriptores visuales para representar el contenido de las imágenes (e.g., presencia de caras, interior/exterior, esquinas, texto, etc.)
- 3 Implementar métodos de clasificación (posiblemente multi-etiqueta) que combinen información visual y textual (posiblemente ensambles).
- 4 Evaluación

Proyecto

Deep GP: Extender el uso de programación genética en el aprendizaje de representaciones inspirados en aprendizaje profundo. Se pretende estudiar las ventajas/limitantes de aplicar programación genética en combinación con modelos pre-entrenados (autoencoders o CNNs).

- 1 Identificar el problema a abordar: clasificación de imágenes.
- 2 Obtener arquitectura pre entrenada y determinar las capas de las redes que podrían ser mejoradas.
- 3 Usar/extender librería de programación genética para combinarla con las redes pre entrenadas.
- 4 Evaluación de los métodos desarrollados.

L. Rodríguez, A. Morales, H. J. Escalante. Evolving Autoencoding Structures through Genetic Programming. Genetic Programming and Evolvable Machines, 2019

L. Rodríguez, A. Morales Reyes, H. J. Escalante. Convolutional Genetic Programming. MCPR, 2019

Proyecto

Aprendizaje profundo localmente ponderado: Extender algún modelo de aprendizaje profundo para operar en un esquema de aprendizaje local.

- 1 Identificar modelos y tareas en los que la adaptación pudiese tener más impacto
- 2 Determinar el esquema de aprendizaje local
- 3 Evaluación

C. G. Atkeson et al. *Locally Weighted Learning*. *Artificial Intelligence Review*. Volume 11, Issue 1-5, pp 11-73, 1997
P. Zadeh et al. *Deep-RBF Networks Revisited: Robust Classification with Rejection*, ArXiv, 2018
Wen-Sheng Chu, Fernando De la Torre, and Jeffrey F. Cohn. *Selective Transfer Machine for Personalized Facial Expression Analysis*. TPAMI, 2017

Proyecto

Word embeddings: Aprendizaje de representaciones de documentos con programación genética a partir de word embeddings.

- Dada una representación distribuida de palabras, aprender a generar la representación de un documento.
- Se sugiere usar programación genética para aprender una función que combine las representaciones de palabras que aparezcan en el documento
- Se espera mejorar el desempeño de representaciones triviales (e.g., promedio) y/o aprendidas (e.g., doc2vec)

Hugo Jair Escalante et al. Term-weighting learning via genetic programming for text classification. *Knowl.-Based Syst.* 83: 176-189 (2015) Distributed Representations of Sentences and Documents Quoc V. Le, Tomas Mikolov, ArXiv, 2014.

Proyecto

AutoCV / AutoText: AutoML en el contexto de clasificación de imágenes o minería de textos

- Diseñar un método automático para resolver problemas asociados con datos visuales o textuales
- Optimizar una arquitectura de aprendizaje profundo, de manera eficiente
- Evaluación online, datos y protocolo de evaluación ya se tiene

Zhengying Liu et al. AutoCV Challenge Design and Baseline Results. Conférence sur l'Apprentissage Automatique, CAP 2019

J. G. Madrid and H.J. Escalante. Meta learning of textual representations. In preparation for AutoML@ICML2019 Workshop, 2019

<http://autodl.chalearn.org>

Proyecto

Clasificación anticipada de documentos/videos:

Reconocer la categoría de una instancia con información parcial

- Seleccionar conjunto de datos (visual o textual), revisión de literatura
- Diseñar método para clasificación anticipada
- Evaluación

P. Lopez, T. Solorio, H. J. Escalante, F. Gonzalez, M. Montes. Early Text Classification using Multi-Resolution Concept Representations. NAACL 2018

H. J. Escalante, et al. Early detection of deception and aggressiveness using profile-based representations. Expert Syst. Appl. 89: 99-111 (2017).

Proyecto

Aprendizaje de embeddings multimodales: Usando embeddings de palabras y una colección de imágenes etiquetada, aprender un espacio embebido multimodal que facilite la recuperación cross-modal de imágenes

- Seleccionar base de datos, definir tarea a abordar
- Estudio de embeddings, avances en etiquetado de imágenes
- Diseño de método de generación de embeddings multimodales
- Evaluación

L. Pellegrin, H. J. Escalante, et al. Overview of the 2017 RedICA Text-Image Matching (RICATIM) Challenge. IEEE-ROPEC 2017 K. Wang, Q. Yin, W. Wang, S. Wu, L. Wang. A Comprehensive Survey on Cross-modal Retrieval. ArXiv, 2016

Proyecto opcional

Challenge: Seleccionar y participar en alguno de las competencias abiertas en [NeurIPS2019](#)

- ① Únicamente si las fechas de la competencia se emparejan con las del curso
- ② Enfocarse en un problema de investigación

<https://neurips.cc/Conferences/2019/CallForCompetitions>

Proyecto ideas generales

- GANs
- Meta-Learning, AutoML
- Neural Architecture Search
- Research on challenge organization (new protocols, challenges for teaching)
- Transfer learning, online learning, meta-learning in evolutionary computation
- Anything on: language & vision

Seguimiento y evaluación (1)

- Propuesta de proyecto (presentación de 10-15 minutos + 5 de preguntas) **Mayo 30**
 - Definir problema, justificar relevancia, describir posibles conjuntos de datos a utilizar, bosquejo preliminar de la solución, técnicas que se utilizarían, identificación de *hueco* (contribución)
- Evaluación de avances (sesiones necesarias, tiempo asignado por determinarse)
 - Sesiones distribuidas durante el curso, el objetivo es dar seguimiento, retroalimentación al proyecto y medir el grado de avance del mismo

Seguimiento y evaluación (2)

- Presentación final (presentación de 20-25 minutos + 5 de preguntas)
 - Presentación general del trabajo (tipo: mini-defensa de tesis / presentación de artículo)
 - Entregar reporte después de presentación de proyecto (formato de artículo tipo LNCS en inglés)
 - Posiblemente se invite a otros profesores a la sesión de entrega

Ideal ... pero no crítico

Que el resultado del trabajo (más algún esfuerzo adicional) pueda someterse y eventualmente publicarse. Algunas opciones:

- Top conference! (ideal, pero complicado, más no imposible): CVPR, ICCV, ECCV, BMVC, ICPR, ACL, EMNLP, NAACL, AAI, IJCAI, GECCO, CEC, EvoStar
- [LatinX@NeurIPS 2019 \(TBC\)](#) Dec. 2019
- [NeurIPS workshops](#) Dec. 2019
- [ROPEC 2019](#) Nov. 2019
- ArXiv
- ?

Preguntas?

