



ME-3008

Procesamiento de Lenguaje Natural y Agentes de Inteligencia Artificial

Duración

Cuatro semanas

Descripción del Curso

Este curso ofrece una formación integral y rigurosa en Procesamiento de Lenguaje Natural (NLP), abarcando desde los modelos vectoriales clásicos del lenguaje – como TF-IDF y análisis semántico – hasta las técnicas modernas de aprendizaje profundo basadas en redes neuronales, embeddings, CNNs, RNNs, LSTMs y Transformers. A lo largo del curso, el estudiantado comprenderá cómo las máquinas representan, procesan y generan lenguaje humano, y cómo estas representaciones se utilizan en aplicaciones reales como modelos de lenguaje de gran escala, extracción de información, grafos de conocimiento y motores de diálogo. El curso culmina con un enfoque práctico y aplicado, donde se diseña e implementa un chatbot inteligente utilizando TensorFlow y arquitecturas Transformer, incluyendo el preprocesamiento de datos, tokenización, mecanismos de atención y atención multi-cabeza, integrando fundamentos matemáticos, estadísticos y algorítmicos con soluciones reales de la industria.

Objetivos

1. **Comprender los fundamentos teóricos del Procesamiento de Lenguaje Natural**, incluyendo modelos vectoriales del lenguaje, representaciones basadas en conteos y análisis semántico, como base para el tratamiento computacional del lenguaje natural.
2. **Analizar y aplicar modelos de aprendizaje profundo para texto**, tales como embeddings, redes neuronales, CNNs, RNNs, LSTMs y arquitecturas Transformer, entendiendo su funcionamiento matemático y algorítmico.
3. **Diseñar e implementar soluciones prácticas de NLP**, integrando técnicas modernas para el desarrollo de aplicaciones reales como modelos de lenguaje, extracción de información, grafos de conocimiento y sistemas conversacionales.



Programa Iberoamericano de
Formación en Minería de Datos

4. **Construir e implementar un chatbot inteligente con TensorFlow**, dominando el flujo completo del proyecto: preparación de datos, tokenización, mecanismos de atención, atención multi-cabeza y despliegue del sistema.

Metodología

El curso combinará teoría y práctica aplicada en Python, permitiendo a los estudiantes crear agentes de Inteligencia Artificial. Las herramientas y recursos incluyen:

1. **Videoconferencias semanales:**
Sesiones en Zoom que serán grabadas para su revisión posterior.
2. **Trabajos prácticos:**
Ejercicios aplicados en Python para aplicar el Procesamiento de Lenguaje Natural en problemas de la industria.
3. **Foros de discusión:**
Espacio en Moodle para la resolución de dudas y discusión de temas del curso.
4. **Aula virtual en Moodle:**
Acceso a materiales, ejemplos de código y recursos adicionales.

Luego de este curso el estudiante será capaz de:

1. Representar texto mediante modelos vectoriales del lenguaje, construir matrices basadas en conteos (TF, TF-IDF) y aplicar técnicas de análisis semántico para el procesamiento computacional de datos textuales reales.
2. Entrenar, evaluar y comparar modelos de aprendizaje profundo para texto, utilizando embeddings, redes neuronales, CNNs, RNNs, LSTMs y arquitecturas Transformer, interpretando su comportamiento desde un punto de vista matemático y algorítmico.
3. Diseñar e implementar aplicaciones completas de Procesamiento de Lenguaje Natural, tales como modelos de lenguaje, sistemas de extracción de información, grafos de conocimiento y motores de diálogo, integrando técnicas modernas de PLN en contextos reales.
4. Construir, entrenar y desplegar un chatbot inteligente con TensorFlow, realizando el preprocesamiento de datos, la tokenización, la implementación de mecanismos de atención y atención multi-cabeza, y la integración final del sistema.

Contenido

- **Semana 1:**
Máquinas de palabras: Modelos vectoriales del lenguaje natural
 - a. Máquinas que leen y escriben: una visión general del procesamiento del lenguaje natural (Natural Language Processing – NLP).

- b. Tokens del pensamiento: palabras del lenguaje natural.
- c. Matemáticas con palabras: vectores de frecuencia de término–frecuencia inversa de documento (TF-IDF).
- d. Encontrar significado en los conteos de palabras: análisis semántico.
- e. Programación en Python de los temas anteriores.
- **Semana 2:**
Aprendizaje profundo (Deep Learning): Redes neuronales
 - a. Cerebro de palabras: redes neuronales.
 - b. Razonamiento con incrustaciones (embeddings) de palabras.
 - c. Encontrar núcleos de conocimiento en texto con CNNs.
 - d. Reducir, reutilizar y reciclar tus palabras: RNNs y LSTMs.
 - e. Implementación en Python.
- **Semana 3:**
Pasarse a lo real: Aplicaciones de procesamiento del lenguaje natural (Natural Language Processing – NLP) en el mundo real
 - a. Aprendizaje profundo apilable: Transformers.
 - b. Modelos de lenguaje grandes en el mundo real.
 - c. Extracción de información y grafos de conocimiento.
 - d. Conversar sin parar: motores de diálogo.
- **Semana 4:**
Construcción de un chatbot con TensorFlow y tecnología Transformer
 - a. Chatbot de IA en pocas palabras
 - ¿Qué es un chatbot?
 - ¿Qué es una palabra de activación (wake word) en un chatbot?
 - Componentes de NLP en un chatbot
 - b. Construcción de un chatbot de películas usando TensorFlow y tecnología Transformer
 - El conjunto de datos del chatbot
 - Preprocesamiento de diálogos de películas
 - Tokenización de conversaciones de películas
 - Proceso de filtrado y padding

- Creación del objeto del conjunto de datos de películas en TensorFlow (mDS)
- Cálculo de pesos de aprendizaje de atención
- Atención multi-cabeza (Multi-Head Attention, MHAttention)

c. Implementación del sistema en Python.

Evaluación

La evaluación consistirá en cuatro entregas prácticas.

Bibliografía

Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd ed.). Prentice Hall.

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press.

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.

Goldberg, Y. (2017). *Neural Network Methods for Natural Language Processing*. Morgan & Claypool Publishers.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Tunstall, L., von Werra, L., & Wolf, T. (2022). *Natural Language Processing with Transformers*. O'Reilly Media.

Rothman, D. (2021). *Transformers for Natural Language Processing* (2nd ed.). Packt Publishing.

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow* (3rd ed.). O'Reilly Media.

Bengfort, B., Bilbro, R., & Ojeda, T. (2018). *Applied Text Analysis with Python*. O'Reilly Media.

Aggarwal, C. C. (2018). *Machine Learning for Text*. Springer.



Programa Iberoamericano de
Formación en Minería de Datos