+33 (0)1 48 10 09 46
formation@ambient-it.net
140 Avenue Jean Lolive 93500 Pantin - France

Updated 03/15/2024

Sign up

Introduction to Deep Learning for Language Processing course

2 days (14 hours)

Presentation

Deep Learning has revolutionized a large number of approaches to language: translation, feature identification, interpretation, dialog systems, and so on. Nevertheless, these new practices require a certain knowledge of the tools, practices and limitations specific to this approach. This course aims to provide the technical keys to implementing and using these new techniques.

The first focus is on the neural network architectures mainly used in this field, and the different implementations and modes of use of these tools. All the main NLP (Natural Language Processing) tasks are covered, with the solutions currently available for each. State-of-the-art models are discussed in detail in a second section, followed by an overview of solutions to date for interpreting or correcting a neural network-based model.

Objectives

- Understand in detail the fundamental architectures of deep learning applied to language processing
- Master the various implementations and tools

Target audience

Developers, Architects, Big Data Data Analyst / Data Engineer / Data Scientist

Prerequisites

- Knowledge of Python and mathematics

Further information

- As an introduction to Artificial Intelligence, we offer you the following training course
- Complementary technology
  - TensorFlow from Google
  - Pytorch from Facebook

Training program Introduction to Deep Learning for language processing

[DAY 1]

1. Introduction Deep Learning

- Reminder of basic mathematics.
- Definition of a neural network: classical architecture, activation and weighting functions for previous activations, network depth.
- Defining neural network learning: cost functions, backpropagation, stochastic gradient descent, maximum likelihood.
- Modeling a neural network: modeling input and output data according to the type of problem (regression, classification, etc.). Curse of dimensionality. Distinction between multi-feature data and signal. Choosing a cost function according to the data.
- Generalization of neural network results.
- Neural network initialization and regularization: L1/L2 regularization, Batch Normalization, Instance Normalization
- Optimization and convergence algorithms

2. Generating embeddings as text templates

- Definition of embedding: unsupervised vocabulary transformation with semantic conservation.
- Word2Vec (skip-gram) approach: generation of word- and sentence-level representations or paragraph.
- FastText framework (Facebook): approaches and use in sentiment detection.
- Direct applications, language modeling. Vocabulary specifics and limitations.

3. Fundamental NLP architectures: recurrent networks

- Introducing Recurrent Neural Networks: fundamental principles and applications.
- Basic operation of RNN: hidden activation, back propagation through time, unfolded version.
- LSTM (Long Short Term Memory). Architecture evolution, gradient flow. Bi-LSTM approach.
- RNN Encoder Decoder architecture.
- Basic NLP applications (sentiment detection, classification).
- Skip Connections and residual construction.
- Study of the Google Neural Machine Translation approach and design. Parallelization, skip-connections.

4. Fundamental NLP architectures: convolutional networks

- Presentation of the fundamental architecture of a CNN layer: convolution, stride, pooling.
- Common non-linearity functions.
- Using pooling.
- Classification issues: cost functions, probabilistic approach.
- Basic NLP applications

5. Attention models and comparisons between convolutional and recurrent approaches

- Attention model for a recurrent network.
- Using a CNN Sequence to Sequence with an attention model. Choice of architecture, implementation example. Application: machine translation.
- Comparisons between recurrent and convolutional approaches. Considerations in quality and performance costs according to problems and approaches.

[DAY 2]

6. Review of language processing applications and architectures: definitions, architectures and results

- Part of speech tagging: Bidirectional LSTMs, Conditional Random Fields.
- Dependency analysis and structuring (Parsing): Stack LSTMs.
- Named entity recognition: Bidirectional LSTM, Dilated CNN and CRFs.
- Sentiment detection: Tree-LSTM with improved embeddings.
- Answering free questions and understanding: Memory Networks, QANet (local convolutions and global self-attention).
- Dialogue system: Dual LSTM Encoder with semantic matching, Sentence-level CNN-LSTM Encoder.

7. Recent approaches and state of the art

- Google's "Attention is all you need" approach
- Transformer cell: a new Deep Learning cell for sequence management. Architecture details, review of an implementation.
- OpenAI ELMO: description of the architecture and advances.
- Description of the BERT model (Google): state of the art and Transfer Learning for NLP. Examples of use.a

8. Visualizing, securing and interpreting a model: tools

- Analysis of RNN cell operation. Activation maxima.
- Visualize attention modes, distinguish dependencies between input and output elements.
- Havard NLP Seq2Seq-Viz tool for visualizing how a model works. Case studies interpretation errors.
- Hierarchical Contextual Decompositions.
- Adversarial attacks: ease of confusion in a neural network. Study of the specific case of language.

9. Free or conditioned text generation

- Presentation of the Variational AutoEncoder architecture: autoencoder evolution, probabilistic approach, reparameterization trick.
- VAE application for text generation.
- Presentation of the GAN (Generative Adversarial Networks) approach: general architecture, convergence principle
- GAN applications to text generation and limitations. Internal metrics issues (Earth Mover Distance)
- Presentation of the Deep Reinforcement Learning approach and applications to text generation.

References : - Generating Sentences from a Continuous Space, Bowman et al. - Toward Controlled Generation of Text, Hu et al.

Companies concerned

This training course is aimed at both individuals and companies, large or small, wishing to train their teams in a new advanced computer technology, or to acquire specific business knowledge or modern methods.

Positioning on entry to training

Positioning at the start of training complies with Qualiopi quality criteria. As soon as registration is finalized, the learner receives a self-assessment questionnaire which enables us to assess his or her estimated level of proficiency in different types of technology, as well as his or her expectations and personal objectives for the training to come, within the limits imposed by the selected format. This questionnaire also enables us to anticipate any connection or security difficulties within the company (intra-company or virtual classroom) which could be problematic for the follow-up and smooth running of the training session.

Teaching methods

Practical course: 60% Practical, 40% Theory. Training material distributed in digital format to all participants.

Organization

The course alternates theoretical input from the trainer, supported by examples, with brainstorming sessions and group work.

Validation

At the end of the session, a multiple-choice questionnaire verifies the correct acquisition of skills.

Sanction

A certificate will be issued to each trainee who completes the course.