

Journée de fin d'année de l'équipe PASTIS

30 juin 2022 — salle A162

14h–14h30 : Accueil et café

14h30–15h30 : Session stagiaires

Chaima Ben Salem

MetaMoi : utilisation des bones pour l'animation de modèles 3D

Il s'agit ici de présenter le contexte du projet MetaMoi et de donner des éléments sur le fonctionnement des modèles 3D animés par Bones. L'objectif actuel est d'implémenter en C/GLSL la fonctionnalité d'animation par Bones en utilisant les bibliothèques Assimp pour l'importation de modèles et GL4D/OpenGL pour la modélisation et le rendu.

Emil Răducanu

Litte_Bot, a chatbot to bring Molière's characters back to life

As we celebrate the quadricentennial anniversary of Molière's birth (1622-1673), we would like to implement an interface where the user will be able to exchange with the characters created by the dramatist. This chatbot should fit each characters' personality, whether in a prose or a verse style. The dataset composed of 32 plays provided by the Bibliothèque nationale de France (BnF) has been preprocessed. We separated prose from verse dialogue in order to train efficiently our models. A dictionary of rhymes coming from Molière's work has been formed as well. We are currently fine tuning a French pretrained multi-layer bidirectional Transformer encoder model, camemBERT, to learn the way of writing of Molière. This knowledge will be then distilled to a Seq2Seq model to finally generate text.

Jilin He

Domain-specific keyword extraction

Keyword extraction is a popular research field in natural language processing, which refers to picking up the terms that are most relevant in a given context. However, most of the datasets are scientific articles which makes the model only good at extracting technological keywords. In this internship, we propose to create a domain-specific keyword extraction model based on BERT to grasp the knowledge from a non-technological field. To do so, we've used thousands of scraped text of companies of different sectors that mention their innovation activities. Also, we used a dictionary of 300 keywords created by Cortext and validated by experst, in order to build our dataset. Finally, the keyword extraction based on BERT model is performed. The results give us new and acceptable discovered keywords.

Mouna Benabid

Reconnaissance Blind Chess

Nous avons pu observer ces dernières années de grandes avancées dans la résolution de jeux complexes tels que le jeu de Go, les échecs, et les jeux Atari, et ce à l'aide de l'apprentissage par renforcement (i.e. Renforcement Learning abrégé RL), qui est un sous-domaine de l'apprentissage automatique (i.e. Machine Learning abrégé ML). Cependant, de telles performances nécessitent des programmes d'avoir une connaissance parfaite de leur environnement. Cette connaissance parfaite est toutefois souvent irréaliste dans les cas réels, où nous n'avons généralement accès qu'à une partie de l'information. Les objectifs de ce travail consistent à étudier l'état de l'art des solutions existantes pour les jeux à information complète et à information incomplète et implémenter des combinaisons de recherche arborescente et de RL en s'adaptant au degré d'information. Les jeux à information incomplète avec reconnaissance sont des jeux à information incomplète où on peut faire des perceptions. Sans perception, il s'agit d'un jeu à information incomplète sans reconnaissance et dans ce cas chaque joueur ne voit que ses pièces. Un exemple de jeu à information incomplète avec reconnaissance est le Reconnaissance Blind Chess (RBC). Dans un premier temps, il s'agit d'établir la liste des algorithmes existants déjà expérimentés sur des jeux à information complète, sur des jeux avec chance, ou sur des jeux à information incomplète. Puis, dans un deuxième temps, d'utiliser certains de ces algorithmes pour proposer une solution pour le jeu RBC.

15h30–15h45 : Pause

15h45–17h15 : Session doctorant·es et ATER

Jean-Pascal Palus

La confiance comme posture de crédulité

La « confiance » (« *trust* »), souvent utilisée de manière interchangeable avec la « conviction » (« *confidence* »), est une position doxastique décrite de multiples façons par les diverses disciplines qui ont tenté de la formaliser. Parfois décrite comme la position épistémique consistant à faire confiance à un agent ou à une proposition, parfois décrite comme l'état psychologique de celui qui fait confiance ; ce concept est néanmoins presque exclusivement décrit comme un moyen d'évaluer un degré de relation entre deux ou plusieurs entités (truster-trustees, truster-proposition) d'un point de vue objectif, ce qui se traduit par l'amalgame de notions telles que la « vérité absolue » ou la « fiabilité » dans ce qui devrait, selon nous, rester un modèle descriptif et prédictif d'adhésion et de confiance dans la valeur de vérité subjective d'une proposition.

Nous présenterons ici un aperçu transdisciplinaire de ce qui se dit sur le sujet dans la littérature et proposerons une approche basée non pas sur l'évaluation externe d'un degré de « confiance » mais sur la modélisation des processus mentaux internes de celui qui fait confiance.

Maroua Boudabous

Active Learning for Explicit Aspect Term Extraction for Low-resource

Language: A use case of French online reviews

We propose to use active learning for an end-to-end process to improve explicit aspect extraction, for low-resource language fine-grained labeling, in opinion and sentiment analysis. With the rapid growth of online commerce, capturing customers' opinions about service and product preferences is essential for business companies to ensure competitiveness and enhance their marketing campaigns. Identifying aspect terms in opinionated texts such as online reviews is crucial to achieving this goal. This task, known as Aspect Term Extraction (ATE) in Natural Language Processing (NLP), remains difficult, especially for low-resource languages where labeled data are lacking. Recent advances in cross-domain learning and active learning make it possible to get sufficiently large datasets and reduce the efforts to annotate them. For the training, we use a web-scraped dataset of French reviews on beauty products and electronic devices. We follow a 3-step process starting from defining a pseudo-labeler CRF using cross-domain learning. Then, we proceed to sequence labelling with BiLSTM-CNN-CRF, deep learning model. Finally, we use active learning to deal with label uncertainty resulting from the first pseudo-labeling step. We adapt the query strategy to fit the sequence labeling setting and handle class imbalance. Results show that active learning significantly improves the learning model's performance achieving over 30% of corrections to the initial labels.

Syrine Saidi

Possibilistic preference networks toolbox

π -pref nets are qualitative models for representing conditional preferences in a possibilistic framework. Preferences of a user are described in terms of degrees of satisfaction.

This software permits to visualize π -pref nets generated from CP-nets. It allows to compare orderings induced from π -pref nets, CP-nets, the Pareto semantic and default rules. It also makes refinements between different orderings.

Oumaima El Joubari

A TDMA-based protocol for safety-related applications in VANET

Vehicular Adhoc Network (VANET) is known as an instantiation of MANET designed to provide ubiquitous connectivity to all road network equipment, in which vehicles use wireless transceivers to communicate with vehicles within their radio communication range. Due to the significance of road safety, one of the stringent requirements of VANET performance is low latency and reliable communication between transport equipment. The frequent changes in the topology, bidirectional mobility, and the continuously varying density can lead to medium access collisions and thus affect the network performance. In a VANET, collisions occur when multiple nodes choose a common physical channel to transmit signals simultaneously specifically due to the hidden terminal problem or the merging collision. To address this issue, a novel channel access mechanism called TA-TDMA was designed to avoid collision issues while minimizing access delays. The protocol aims to mitigate the unfairness in time resources allocation by utilizing a traffic prediction model. Traffic conditions on opposite directions, different lanes, and road segments are forecasted and then each frame is partitioned accordingly. Each direction, lane, and segment is allocated a different set of time slots while considering the vehicle traffic conditions. The protocol is also cluster-based and access time among participating nodes is managed by the cluster head (CH).

17h15–18h : Convivialité