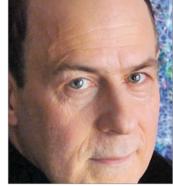


DOCTORAT EN INFORMATIQUE COGNITIVE



Claude Coulombe

Techniques d'amplification des données textuelles pour l'apprentissage profond

25 octobre 2019, 13 h 30, Université TÉLUQ 5800, rue Saint-Denis, Montréal, local 11.051 (amphithéâtre) Webdiffusion en direct : conf.teluq.ca/soutenance/

Cette thèse a pour objectif d'étudier la faisabilité de différentes techniques d'amplification des données textuelles (ADT) afin de contrer l'insuffisance de données pour entraîner de gros modèles statistiques, particulièrement en apprentissage profond.

Pour franchir ce « mur des données massives », nous proposons de créer de nouvelles données à partir des données existantes. On dit aussi : données synthétiques, données générées ou données artificielles.

Nous recherchons des techniques d'ADT pratiques, robustes, capables de monter en charge et simples à mettre en œuvre. Ces techniques de prétraitement s'inspirent des techniques utilisées avec succès en vision artificielle.

Aux fins de comparaison, des techniques d'ADT existantes ont été expérimentées, comme l'injection de bruit ou l'emploi d'expressions régulières. D'autres ont été améliorées, comme la substitution lexicale. Enfin, des techniques plus innovatrices utilisent des services en ligne comme la rétrotraduction et la génération de paraphrases par transformation d'arbres syntaxiques.

Les techniques d'ADT étudiées ont permis d'accroître l'exactitude de 0,5 à 8,8 %, sur une tâche de prédiction de la polarité de textes. Différentes architectures de réseau de neurones ont été testées : le perceptron multicouche (PMC), le réseau convolutif 1D (RNC 1D), le réseau récurrent à longue mémoire court terme (LMCT) et le réseau récurrent LMCT bidirectionnel (biLMCT).

JURY D'ÉVALUATION

GILBERT PAQUETTE

Professeur

Département Science et Technologie, Université TÉLUQ

NEILA MEZGHANI

Professeure

Département Science et Technologie, Université TÉLUQ

MICHEL GAGNON

Professeur

Polytechnique Montréal

BELKACEM CHIKHAOUI

Professeur

Département Science et Technologie, Université TÉLUQ

STAN MATWIN

Professeur

Faculty of Computer Science, Université Dalhousie

CHRISTOPHE MALATERRE

Professeur

Département de philosophie, UQAM

PRÉSIDENT DU JURY

BELKACEM CHIKHAOUI

DIRECTION DE RECHERCHE

GILBERT PAQUETTE

Directeur de recherche

NEILA MEZGHANI

Codirectrice de recherche

MICHEL GAGNON

Codirecteur de recherche



