

Machine Learning on Big Data [MLBD]

Responsable	Pena Carlos Andrés
MRU	TIC / HEIG-VD
Profils/Options concerné-e-s	TIC / Réseaux de télécommunications et sécurité de l'information TIC / Système d'informations complexes
Contrainte temporelle	Semestre 2
Capacité minimum	5
Capacité maximum	25
Groupe de site	Lausanne
Résumé	Prof. Carlos Peña (HEIG-VD) http://iict-space.heig-vd.ch/cpn Prof. Andres Perez-Urbe (HEIG-VD) http://iict-space.heig-vd.ch/ape

Avec l'essor de l'informatique ubiquitaire, on entre dans l'ère des Big Data, des ensembles de données tellement gigantesques qu'ils nécessitent de nouveaux outils techniques et scientifiques pour les comprendre et en tirer du sens. L'objectif de ce cours est d'appliquer une série de techniques de Machine Learning (Deep networks, Neural networks, Fuzzy logic, Evolutionary Fuzzy Modeling, Random Forests, Self-Organizing Maps, Growing Neural Gas, etc.) à des données issues de la bioinformatique, de capteurs biotechnologiques, de capteurs des smartphones, des infrastructures de télédétection (remote sensors), etc., et d'analyser les résultats dans le but de produire de la connaissance à partir d'un "déluge de données".

Ce cours présente une approche pratique sur le pré-traitement, la modélisation, l'analyse et la visualisation de grands ensembles de données, de données à haute dimensionnalité, et de données bruitées, incomplètes ou non numériques. Nous tirerons profit des algorithmes d'apprentissage automatique bio-inspirés et utiliserons des bibliothèques parallélisables facilitant le traitement, la modélisation et l'analyse de ces données.

A la fin de ce cours, l'étudiant-e est capable de:

*Argumenter, expliquer et justifier le choix d'une technique de Machine Learning pour construire des modèles avec une approche "data-driven"

*Choisir les techniques de machine learning en fonction du domaine

*Analyser les résultats de modèles répondant à diverses configurations et paramètres selon les domaines d'application

*Analyser les avantages et désavantages de l'utilisation de modèles issus de différentes approches

*Utiliser des bibliothèques parallélisables facilitant le traitement, la modélisation et l'analyse de grandes quantités de données

*Tirer profit de l'expérience acquise au niveau du traitement de données dans les domaines d'application traités dans le cours

Machine Learning on Big Data [MLBD]

Contenu

Sujet	Temps [%]
Machine Learning: changement de paradigme et challenges dans l'ère des Big Data	7
Machine learning dans le domaine du traitement d'images	7
Classification, reconnaissance et similarité	7
Machine learning dans le traitement de données biologiques	14
Ensembles de données à (très) haute dimensionnalité	7
Machine learning dans le développement de tests diagnostiques	7
Sélection de modèles: performance et interprétabilité	7
Machine learning dans le domaine du traitement de séries temporelles	7
Reconnaissance de patterns et détection d'anomalies	7
Conclusions et perspectives	2
Mini-projet	28

Connaissances préalables

Bonnes bases de programmation, Java, C, C++
Techniques de Machine Learning

Modules pré-requis

Aucun

MA's exclusifs

Aucun

Méthodes d'enseignement

Mode	Périodes d'enseignement	Volume de travail (en heures)
Exposés	18	38.57
Exercices	0	0
Travaux pratiques	24	51.43
TOTAL	42	90
Crédits ECTS		3

Évaluation (2016-2017)

Examen oral : 70%
Contrôle(s) continu(s) / projet(s) : 30%

Langues

	Français	Allemand	Anglais	Italien
Enseignement	X			
Documentation	X		X	
Questions d'examen	X			

Machine Learning on Big Data [MLBD]

Compétences visées

Gérer le projet	25%
Gérer l'avancement technique	
Gérer les coûts et les délais	
Gérer la communication	
Gérer les risques et les imprévus	
Stimuler l'équipe	
Analyser et spécifier des produits / services	25%
Analyser le système (pluridisciplinarité)	
Décomposer le système	
Spécifier le système, y compris concept	
Evaluer les risques	
Planifier	
Développer et réaliser	30%
Analyser et spécifier en détail	
Concevoir	
Intégrer (d'autres composants ou produits)	
Modéliser / simuler	
Mesurer / tester / caractériser	
Documenter (rapport)	20%
Analyser / critiquer	
Proposer les améliorations	
Tirer les leçons / apprendre	
Documenter	
Disséminer	