



AVIS DE SOUTENANCE DE THESE

Le Doyen de la Faculté des Sciences Dhar El Mahraz –Fès – annonce que

Mme (elle) : **KARIM Awatif**

Soutiendra : **le 12/04/2021 à 15H**

Lieu : **Centre de Visioconférence**

Une thèse intitulée :

Modèles intelligents à base des réseaux de neurones pour la classification des documents textuels

En vue d'obtenir le Doctorat

FD : Sciences et Technologies de l'Information et de la Communication (STIC)

Spécialité : Informatique

	NOM ET PRENOM	GRADE	ETABLISSEMENT
Président	Pr. EL BEQQALI Omar	PES	Faculté des Sciences Dhar El Mahraz - Fès
Directeur de thèse	Pr. BOUMHIDI Jaouad	PES	Faculté des Sciences Dhar El Mahraz - Fès
Co-directeur de thèse	Pr. LOQMAN Chakir	PH	Faculté des Sciences Dhar El Mahraz - Fès
Rapporteurs	Pr. SABBANE Mohammed	PES	Faculté des Sciences - Meknès
	Pr. NFAOUI El Habib	PES	Faculté des Sciences Dhar El Mahraz - Fès
	Pr. MAJDA Fikri	PH	Ecole Nationale de commerce et de Gestion - Agadir
Membres	Pr. HADDOUCH Khalid	PH	Ecole Nationale des Sciences Appliquées - Fès
	Pr. SOSSE ALAOUI Mohammed Chakib	PH	Centre Régional des Métiers de l'Education et de la Formation - Fès

Résumé :

Le nombre de données non structurées forme environ 90% de données disponible sur le web et sur les supports de stockage. Or la gestion de ce nombre énorme de documents numériques, qui ne cesse d'évoluer à chaque instant, nécessite les techniques d'intelligence artificielle pour pouvoir classifier et gérer automatiquement ces données de haute dimension. Nous proposons des nouveaux modèles intelligents à base des réseaux de neurones artificiels.

Nous avons utilisé dans la première contribution de la présente thèse, d'une part la théorie des graphes; précisément le concept de l'ensemble stable maximum (MSSP) pour modéliser le problème de clustering de texte, et d'autre part les réseaux de Hopfield continu comme réseaux de neurones, pour détecter automatiquement le nombre de clusters et les centres initiaux du corpus proposé. Ces derniers seront les paramètres de base de K-Means dans la deuxième contribution. Notre approche a prouvé son efficacité et sa performance en terme de qualité de clustering et de temps d'exécution pour les grands ensembles de données.

Dans la dernière partie, nous avons réalisé une synthèse sur la structure et les implémentations techniques des systèmes de classification de textes basés sur l'apprentissage profond. L'objectif est d'examiner l'impact de nombreuses représentations de mots traditionnelle par rapport au plongement de mots contextuelle (BERT) sur la réalisation de classification de textes.

Mots clés :

Apprentissage non supervisé, Analyse de données, Classification, Document clustering, Text mining, Nombre de clusters, K-Means, Ensemble stable maximum, Réseaux de neurones, Réseau de Hopfield, Apprentissage profond, Représentation de texte.

Intelligent models based on neural networks for the classification of textual documents

Abstract :

The number of unstructured data forms about 90% of the data available on the web and on storage media. Managing this huge amount of digital documents, which is constantly evolving, requires artificial intelligence techniques to automatically classify and manage this high-dimensional data. We propose new intelligent models based on artificial neural networks.

In the first contribution of this thesis, we used graph theory; specifically, the concept of maximum stable set (MSSP) to model the text clustering problem, and continuous Hopfield networks as neural networks, to automatically detect the number of clusters and the initial centers of the proposed corpus. These will be the basic parameters of K-Means in the second contribution. Our approach has proven its efficiency and performance in terms of clustering quality and execution time for large data sets.

In the last part, we performed a synthesis of the structure and technical implementations of text classification systems based on deep learning. The objective is to examine the impact of many traditional word representations versus contextual word embedding (BERT) on the realization of text classification.

Key Words: Unsupervised learning, Data analysis, Classification, Document clustering, Text mining, Number of clusters, K-Means, Maximum stable set Problem, Neural networks, Hopfield network, Deep learning, Text representation.