



CNRS - INP - UT3 - UT1 - UT2J

Institut de Recherche en Informatique de Toulouse



EQUIPE SIG

Données, Systèmes d'information Généralisés Apprentissage et Deep Learning

Geoffrey Roman Jimenez, Josiane Mothe



EQUIPE SIG

Systeme d'Informations Généralisées

- Structure
 - 19 permanents dont 8 PR et 2 MCF HDR
 - 38 non-permanents dont 23 doctorants et 5 post-doc
- Axe de recherche
 - ***De la donnée brute vers une information intelligible***
 - Intégration d'informations hétérogènes
 - Modélisation de l'information
 - Données orientées utilisateurs
 - Traitements d'informations des Big Data

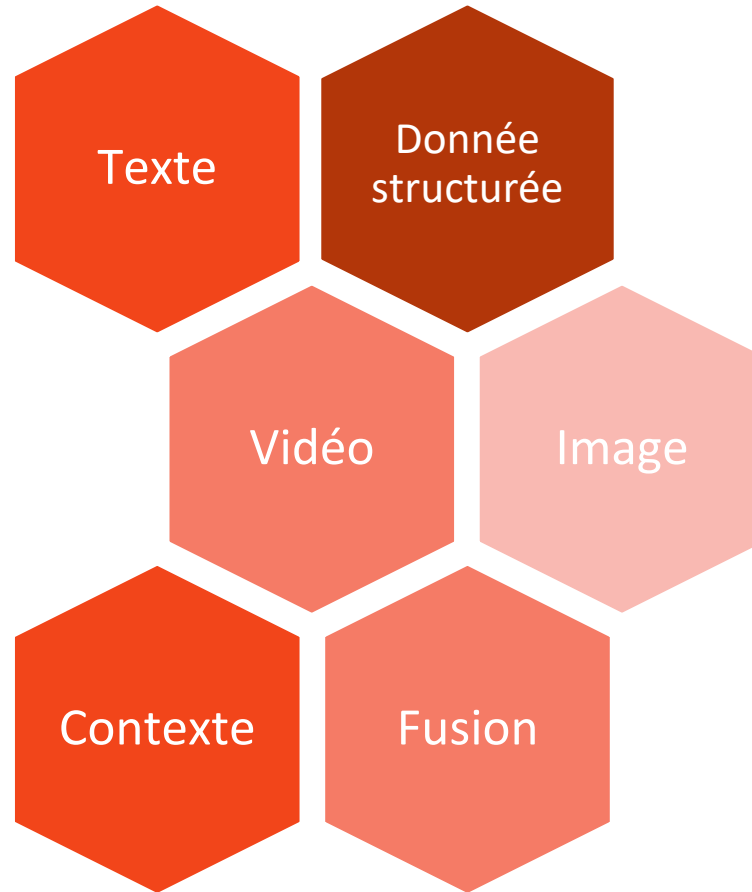
Système d'Informations Généralisées

 **Verrous et axes scientifiques :**

- **Activité (1) Intégration d'informations hétérogènes**
 - Intégration holistique : sources multiples (>2)
 - Intégration de sources massives : Variété, +Volume, +Vélocité
 - Alignement *vs* unification des données
- **Activité (2) Modélisation de l'information**
 - Conception des entrepôts de données (*data warehouses*) multidimensionnelles complexes (textuelles, semi-structurées, agrégations multiples)
 - Meta-modélisation de données spatio-temporelles, multi-médias, vidéos, multi-couches
 - Modélisation des bases de données non-conventionnelles : noSQL, multi-modèles
- **Activité (3) Intégration de données orientées utilisateurs et usages**
 - Système de recommandation
 - » Profils utilisateurs (pertinence, égocentré, social, évolution)
 - » Contextualisation, Diversification
 - » Scénarios et Traces d'usages
 - Recherche d'informations / Requêtes difficiles : désambiguïsation, contextualisation
 - Réseaux sociaux, Internet des objets : Propagation de l'information, Qualité (malveillance, influence dans les données)
- **Activité (4) Analyse et Apprentissage dans les données massives**
 - Fouille de données (« Data Mining »)
 - » Détection & Découverte des phénomènes typiques/atypiques
 - » Meta-mining (Réutilisation, Recommandation, Prédiction)
 - » Visualisation des graphes dynamiques & vastes (temporalité, obsolescence)
 - Apprentissage machine (« Machine Learning »), Apprentissage profond (« Deep Learning »)
 - » Structuration et paramétrisation des données pour améliorer l'apprentissage
 - » Transfert de modèles d'apprentissage



Deep learning sur différents types de données



Représentation des textes Word2vec

Fouille : détection de comportements/changements

Texte

Donnée structurée

Représentation parcimonieuse dans le contexte de vidéo surveillance

Vidéo

Image

Classification et détection de changement (couverture, déforestation)

Classification d'images médicales (multi-classes)

Contextualisation des données

Contexte

Fusion

Fusion de média (texte, image, multi-échelle, multi-variée)



Deep learning & Apprentissage pour différentes applications

- Vidéo surveillance, sécurité, fouille
 - Ré-identification des objets (véhicule/piétons) au cours du temps et dans différents contextes (endroits, jours/nuit, qualités, point de vue, etc.)
 - Fouille de réseaux sociaux (textes semi-structurés) pour détection de comportement (agressivité, ...)
- Santé
 - Classification d'images (multi-classes) pour la détection automatisée de cancer (CNN)
 - Segmentation d'images médicales basée sur des CNN 3D pour la localisation des zones de Traumatisme Crânien et la classification de sa sévérité
 - Détection de déséquilibre de marche chez les personnes âgées : utilisation de données de centrale inertielle d'une canne connectée

Deep learning & Apprentissage

Verrous scientifiques

- Représentation
 - Modélisation de données hétérogènes; fusion de données multi-échelles, multi-médias, multi-variées
 - Parcimonieuses: algorithmes de sélection par apprentissage (SVM parcimonieux...)
 - Big data et Small data
- Apprentissage avec peu de données d'apprentissage
 - Augmentation de données
 - Classes déséquilibrées
 - Apprentissage par le contexte
- Méta-analyse (méta-données des algorithmes, topologie des jeux de données)
 - Recommandation de workflows, prédictions, explication des prédictions
 - Structuration et paramétrisation des données et méta-données pour améliorer l'apprentissage
- Transfert d'apprentissage
 - Réutilisation des connaissances
 - Réduire la dépendance au domaine d'apprentissage



Deep learning & Apprentissage Projets en cours

- Projet OPTIMIS-TC (collaboration CHU-Toulouse)
- Projet CIPAD (M2 financé CIMI, collaboration EHPAD AGIR Castres)
- Collaboration Oncopole, IRIT REVA & SIG
- FabSpace 2.0 (image EO H2020) - **Ce projet a reçu un financement de l'Union Européenne (programme Recherche et Développement Horizon 2020 - convention n°693210).**
- Projet ModelAge (thèse financée par la région et la COMUE, avec I2MC-INSERM)

- Ré-identification des véhicules/piétons en vidéo surveillance par encodage parcimonieux des espaces latents des réseaux profonds
- Toulouse Campus Surveillance Dataset (TOCADA): un Dataset de video-surveillance
- Aide au diagnostic et à la prévention (cancer, dépendance et maladies liées au vieillissement)
- Calculs à partir d'images satellites (déforestation, population*, eau)

* challenge en cours!!

<https://www.irit.fr/eofusion/participation-challenge/>

- *Sonia Mejbri, Aide à la détection des cancers avec l'apprentissage profond*
- *Nathalie Neptune, Détection de la déforestation sur des images satellites avec l'apprentissage profond*
- *Mahmoud Qodseya, Social Interaction Analysis*
- *William Raynaud et Gabriel Ferrettini : Meta-Analysis for Workflow Recommendation*
- *Geoffrey Roman-Jimenez : Représentation parcimonieuse des espaces latents des réseaux profonds pour la re-identification des véhicules/pietons dans un contexte de vidéosurveillance.*
- *Geoffrey Roman Jimenez, Dataset de video-surveillance: Toulouse Campus Surveillance Dataset*



Geoffrey Roman Jimenez

Ré-identification de véhicule/personne en utilisant le signal visuel

