

Les utilisations possibles de l'Intelligence Artificielle dans la linguistique historique

3 étudiants de CPBx

- 0.1 Résumé
- 0.2 Abstract
- 0.3 Remerciements

Table des matières

| | | |
|----------|--|----------|
| 0.1 | Résumé | 1 |
| 0.2 | Abstract | 1 |
| 0.3 | Remerciements | 1 |
| 0.4 | Table des figures | 2 |
| 0.5 | Notations | 2 |
| 1 | Introduction | 3 |
| 2 | La linguistique historique et l'Intelligence Artificielle | 4 |
| 2.1 | La linguistique historique | 4 |
| 2.1.1 | Introduction à la linguistique historique | 4 |
| 2.1.2 | Les différents principes | 4 |
| 2.1.3 | Les atouts de l'Intelligence Artificielle dans ce domaine | 4 |
| 2.2 | L'IA dans le Traitement Automatisé du Langage Naturel | 5 |
| 2.2.1 | Introduction à l'apprentissage automatique | 5 |
| 2.2.2 | Les pré-traitements nécessaires du texte. | 5 |
| 2.2.3 | Architectures neuronales utiles au TAL | 5 |
| 2.2.4 | Tâches fondamentales de TAL réalisées par l'IA | 5 |
| 3 | Les contributions de l'IA dans la linguistique historique | 6 |
| 3.1 | Traitement de texte | 6 |
| 3.1.1 | Numérisation de textes historiques | 6 |
| 3.1.2 | Complétion de ces textes | 6 |
| 3.2 | Analyse de document | 6 |
| 3.2.1 | Extraction de données linguistiques | 6 |
| 3.2.2 | Identification de motifs | 6 |
| 3.2.3 | Classification de ces documents | 6 |
| 3.3 | Modélisation linguistique | 6 |
| 3.3.1 | Conception d'un modèle de langue | 6 |
| 3.3.2 | Simulation vers l'évolution d'une langue | 6 |
| 4 | Utilisation de l'IA pour la reconstruction de la proto-forme d'une langue au travers d'une expérience | 7 |
| 4.1 | Mise en place de l'expérience | 7 |
| 4.1.1 | Conceptualisation du problème | 7 |
| 4.1.2 | Récupération de la base de données | 7 |
| 4.1.3 | Potabiliser les données | 7 |
| 4.2 | Démarche statistique | 7 |
| 4.2.1 | Méthode | 7 |

| | | |
|----------|--|-----------|
| 4.2.2 | Analyse | 7 |
| 4.2.3 | Critiques | 7 |
| 4.3 | Démarche neuronale | 8 |
| 4.3.1 | Méthode | 8 |
| 4.3.2 | Analyse | 8 |
| 4.3.3 | Critiques | 8 |
| 5 | Conclusion | 9 |
| 5.1 | Synthèse | 9 |
| 5.2 | Les différentes limites posées aujourd’hui | 9 |
| 5.3 | Les perspectives de l’IA dans la linguistique historique | 9 |
| 6 | Références | 10 |
| 6.1 | Bibliographie | 10 |

0.4 Table des figures

0.5 Notations

Chapitre 1

Introduction

Mise en contexte pour arriver à la problématique, quel est le potentiel de l'intelligence artificielle dans la linguistique historique ? ... ?

Chapitre 2

La linguistique historique et l'Intelligence Artificielle

2.1 La linguistique historique

2.1.1 Introduction à la linguistique historique

Définir ce qu'est la linguistique historique, ce qu'elle étudie, et les mots de vocabulaires que nous allons rencontrer tout au long du mémoire.

2.1.2 Les différents principes

Évidemment cette science repose sur des concepts, allant des propriétés synchroniques des mots aux à leurs aspects diachroniques.

2.1.3 Les atouts de l'Intelligence Artificielle dans ce domaine

La linguistique historique fait face à de nombreux problèmes récurrents (traiter une grande quantité de textes pour l'homme, remarquer des motifs dans ces documents historiques). Alors que ce travail pourrait être effectué par une machine, grâce à sa capacité à traiter un grand nombres de données, et à chercher des similarités dans ces données. Avant, de voir les tâches où l'Intelligence Artificielle peut intervenir, il est d'abord nécessaire de voir en détail la conception des ces IA.

Résoudre des problèmes de Linguistique Historique avec un ordinateur nécessite de lui faire traiter du contenu textuel devant être abstrait sur des terrains parmi ceux de la **phonétique**, de la **sémantique**, de la **morphologie** ou encore de la **syntaxe**.

Développer un exemple pour illustrer ces 4 niveaux d'abstractions

La réalisation de ces abstractions s'inscrit dans le Traitement Automatisé du Langage Naturel (TAL), un domaine à cheval entre la Linguistique et l'Informatique. L'Intelligence Artificielle y occupe une place centrale pour sa capacité à effectuer des approximations améliorables avec de l'entraînement.

2.2 L'IA dans le Traitement Automatisé du Langage Naturel

2.2.1 Introduction à l'apprentissage automatique

Qu'est ce qu'une intelligence artificielle ?

Qu'est ce qu'un réseau de neurones ?

Quel est le principe derrière l'apprentissage automatique ?

Définition des apprentissages supervisés/non supervisés

2.2.2 Les pré-traitements nécessaires du texte.

tokenisation + normalisation des données

Vectorisation sémantique des mots (embeddings)

Encodage d'embeddings

2.2.3 Architectures neuronales utiles au TAL

Quels sont les différents outils ?

Suivant, comment les parties précédentes ont été traitées, ou comment les parties futures seront discutées, cette partie pourrait ne pas être nécessaire. Sinon, elle regroupera l'idée de comment on passe de notre langue naturelle à celle de la machine, de passer aux mots à des vecteurs ? Quels traitements théoriques (théoriques pour ce distinguer de la pratique dans la partie future) doivent être effectués sur les mots ? En fait cette partie fait référence aux chapitres 2 et 6 de Jurasky. Voir même le chapitre 9, en supprimant la sous partie précédente pour pouvoir parler directement des réseaux de neurones appliqués à la linguistique, en d'autres termes, des réseaux récurrents, des modèles séquentielles (encodeurs-décodeurs) avec l'attention, et des Transformers.

2.2.4 Tâches fondamentales de TAL réalisées par l'IA

blablabla

Chapitre 3

Les contributions de l'IA dans la linguistique historique

C'est la partie 'Related Work', elle discute des différents aspects où la linguistique historique s'applique, à travers différents modèles.

3.1 Traitement de texte

3.1.1 Numérisation de textes historiques

3.1.2 Complétion de ces textes

3.2 Analyse de document

3.2.1 Extraction de données linguistiques

3.2.2 Identification de motifs

3.2.3 Classification de ces documents

3.3 Modélisation linguistique

3.3.1 Conception d'un modèle de langue

3.3.2 Simulation vers l'évolution d'une langue

Chapitre 4

Utilisation de l'IA pour la reconstruction de la proto-forme d'une langue au travers d'une expérience

Ici, on se place dans un cas concret, pour montrer que ce n'est pas que de la théorie. En proposant une expérience.

4.1 Mise en place de l'expérience

4.1.1 Conceptualisation du problème

Définir clairement le problème du titre, énoncer et justifier le choix de notre modèle réseaux de neurones et des différents outils appliqués. Voir s'il est possible de faire apparaître plusieurs démarches, c'est à dire, une approche statistique et une approche neuronale (toujours pour renforcer et montrer le potentiel de l'IA).

4.1.2 Récupération de la base de données

Il est fort possible que cette partie se regroupe avec la partie suivante, car il n'y aura pas grand chose à dire.

4.1.3 Potabiliser les données

Le choix de "potabiliser", et non pas normaliser, est volontaire. En effet, la normalisation de nos données s'effectuera dans un second temps dans les différentes démarches.

4.2 Démarche statistique

4.2.1 Méthode

4.2.2 Analyse

4.2.3 Critiques

Il reste ici quelques sous parties à détailler.

4.3 Démarche neuronale

4.3.1 Méthode

4.3.2 Analyse

4.3.3 Critiques

Il reste ici quelques sous parties à détailler.

Chapitre 5

Conclusion

5.1 Synthèse

Résume tout ce qui a été dit.

5.2 Les différentes limites posées aujourd’hui

une partie des limites aura déjà été traitée dans le chapitre précédent. Cette sous partie se veut résumer ces limites, et aller dans les limites générales (voir acutelles) de l’IA dans la linguistique historique.

5.3 Les perspectives de l’IA dans la linguistique historique

Ouverture, dépassement de certaines limites, évolution des modèles.

Chapitre 6

Références

6.1 Bibliographie