

Détection des effets indésirables d'un médicament à partir d'un texte

Eden BELOUADAH & Mariem BOUHAHA

Introduction

- L'Intelligence artificielle a plusieurs applications dans le domaine médical.
- **But:** développer un système automatique de détection des effets indésirables dans un texte/notice.
- **Hypothèse de base:** utiliser CRF Wapiti avec comme entrée un fichier contenant le mot et sa classe avec un fichier de patrons de base.

Matériel

- Nous avons à notre disposition: 101 fichiers (chacun pour un médicament).
- Format du fichier: tabulaire.

Nous avons divisé les fichiers:

- 60% pour l'apprentissage.
- 20% pour le développement.
- 20% pour le test.

```
AMPYRAtab
2 ADVERSE AMPYRA.xml:S1:6:7 O
3 REACTIONS AMPYRA.xml:S1:14:9 O
4
5 Because AMPYRA.xml:S1:27:7 O
6 clinical AMPYRA.xml:S1:35:8 O
7 studies AMPYRA.xml:S1:44:7 O
8 are AMPYRA.xml:S1:52:3 O
9 conducted AMPYRA.xml:S1:56:9 O
10 under AMPYRA.xml:S1:66:5 O
11 widely AMPYRA.xml:S1:72:6 O
12 varying AMPYRA.xml:S1:79:7 O
13 conditions AMPYRA.xml:S1:87:10 O
14 , AMPYRA.xml:S1:97:1 O
15 adverse AMPYRA.xml:S1:99:7 O
16 reaction AMPYRA.xml:S1:107:8 O
17 rates AMPYRA.xml:S1:116:5 O
18 observed AMPYRA.xml:S1:122:8 O
19 in AMPYRA.xml:S1:131:2 O
20 the AMPYRA.xml:S1:134:3 O
21 clinical AMPYRA.xml:S1:138:8 O
```

Figure 2. Fichier tabulaire pour la notice du médicament AMPYRA

Méthodes et protocole expérimental

Nous avons suivis plusieurs étapes:

1. Hypothèse de base
2. Bigrammes de classes
3. Racinisation
4. Etiquetage morpho-syntaxique
5. Ne pas tenir compte de la casse
6. Lemmatisation
7. Représentation vectorielle
8. Elimination des classes inutiles
9. Elimination des mots vides

Evaluation: F-mesure + token error + sequence error.

Résultats

Les résultats obtenus pour les différentes étapes sur l'ensemble de développement:

Etape	0	1	2	3	4	5	6	7	8
Tokens error	5.59%	5.53%	5.18%	4.94%	4.80%	4.89%	5.36%	3.84%	4.96%
Sequences error	35.59%	30.09%	29.58%	27.33%	26.62%	26.90%	27.07%	23.71%	23.62%
F(B-AdvReaction)	72%	73%	74%	75%	77%	77%	72%	77%	75%
F(I-AdvReaction)	62%	66%	68%	68%	69%	69%	61%	68%	67%

Tableau 2. Résultats obtenus pour les différentes étapes

Les résultats finaux obtenus sur l'ensemble de **test**:

Ensemble	Développement	Test
Tokens error	3.84%	3.25%
Sequences error	23.71%	16.71%
F _{mesure} (B-AdverseReaction)	77%	86%
F _{mesure} (I-AdverseReaction)	68%	74%

Tableau 3. Résultats finaux pour le développement et le test

Discussion des résultats

- Les résultats s'améliorent d'une étapes à une autre,
- Sauf pour certaines étapes que nous avons décidé de ne pas prendre en considération
- Les résultats sur le test sont mieux que ceux sur l'ensemble de développement.

Et si on s'intéresse que
s'il s'agit d'un effet
indésirable ou non?

Ensemble	Développement	Test
Tokens error	3.72%	3.11%
Sequences error	23.18%	16.50%
$F_{mesure}(\text{AdverseReaction})$	76%	82%

Tableau 4. Résultats de l'expérience simplificatrice

Conclusion

- La détection des effets indésirables peut être étendue à d'autres tâches similaires telles la détection des traitements et des symptômes
- L'hypothèse de base a donné un résultat acceptable.
- Mais il fallait l'enrichir avec d'autres informations pertinentes.
- **Perspectives:** Se servir des ressources externes