

nty nty/global/

hajjar 00hajjar

MÉMOIRE

pour obtenir le grade de Master délivré par

Université Paris 8 Vincennes à Saint-Denis

Mention *Informatique*

Parcours MIASHS Big data et fouille de données

présenté et soutenu publiquement par

Salahh Bakir

le 16 septembre 2017

Evaluation des classifications non supervisées

Encadrant universitaire : Jean-Jacques MARIAGE

Université Paris 8
Laboratoire d'Informatique Avancée de Saint Denis
EA n° 4383 Saint Denis, France

Sommaire

Introduction	5
Problématique	7
1 Le contexte de résolution du problème	11
1.1 Introduction	12
1.2 Conclusion	12
2 Le problème à résoudre	13
2.1 Introduction	14
2.2 Conclusion	14
État de l’art	15
3 État de l’art des techniques	19
3.1 Introduction	20
3.2 Indices externes	20
3.2.1 Taux global d’erreur de classification (T_{gec}) -sans coût- . . .	20
3.3 Conclusion	20
4 État de l’art des données	21
4.1 Introduction	22
4.2 Conclusion	22

Système réalisé	23
5 Implémentation du système	27
5.1 Introduction	28
5.2 Conclusion	28
6 Expérimentations et résultats	29
6.1 Introduction	30
6.2 Conclusion	30
Conclusion	31

Introduction

Ce chapitre est à faire à la fin. Il doit mettre en valeur le travail effectué par l'étudiant. Si un lecteur ne lit que l'introduction, il doit avoir envie de lire le reste.

Première partie

Problématique

Sommaire

1	Le contexte de résolution du problème	11
1.1	Introduction	12
1.2	Conclusion	12
2	Le problème à résoudre	13
2.1	Introduction	14
2.2	Conclusion	14

Chapitre 1

Le contexte de résolution du problème

Sommaire

1.1	Introduction	12
1.2	Conclusion	12

1.1 Introduction

Présenter ici les contraintes particulières posées par la société ou le projet.

1.2 Conclusion

Chapitre 2

Le problème à résoudre

Sommaire

2.1	Introduction	14
2.2	Conclusion	14

2.1 Introduction

Il y a souvent deux aspects à la problématique : le problème à résoudre et le contexte dans lequel il doit être résolu (par exemple, le contexte de la société où le stage est fait)

2.2 Conclusion

Deuxième partie

État de l'art

Sommaire

3	État de l'art des techniques	19
3.1	Introduction	20
3.2	Indices externes	20
3.3	Conclusion	20
4	État de l'art des données	21
4.1	Introduction	22
4.2	Conclusion	22

Chapitre 3

État de l'art des techniques

Sommaire

3.1	Introduction	20
3.2	Indices externes	20
3.2.1	Taux global d'erreur de classification (T_{gec}) -sans coût- .	20
3.3	Conclusion	20

3.1 Introduction

Il existe Plusieurs indices pour mesurer la qualité d'une classification non supervisée. Ils sont principalement de deux sortes; les indices externes et les indices internes. L'utilisation des indices externes, requière le recours à des information à priori sur les classes. Tandis que les critères internes, reposent sur la cohérence de la répartition des données dans la classification (similarité, dé-similarité, inertie des classes et cohésion topographique).

Dans cette partie, nous allons passer en revue les principaux indices de qualité. Nous verrons en premier lieu, les indices externes et dans un second les internes. Pour les premiers, nous allons nous intéresser au taux global d'erreur de classification (Overall Error Rate of Classification OERC) sans coût et avec coût, sections 3.1.1 et 3.1.2. En section 3.1.3 nous aborderons les courbes ROC (Receiver Operating Characteristics) et à l'Aire sous courbes ROC (Area Under the Curve AUC) en section 3.1.4. Et enfin nous traiterons les mesures de Rappel, Précision et F-mesure (Precision, Recall and F-measure), respectivement en section 3.1.5, 3.1.6 et 3.1.7. En second lieu, nous nous intéresserons au mesure internes. indice de Dunn (1974) section 3.2.1 indice BH de Baker et Hubert (1975) section 3.2.2 indice DB de Davies et Bouldin (1979) section 3.2.3 indice CH de Calinsky et Harabsz (1974) section 3.2.4 indice Silhouette (S) de Rousseeuw (1987) section 3.2.5 indice HL de Hubert et Levin (1976) section 3.2.6 indice KL de Krzanowski et Lai (1988) section 3.2.7

3.2 Indices externes

3.2.1 Taux global d'erreur de classification (T_{gec}) -sans coût-

Salah dhjklfdjsdl [?].

3.3 Conclusion

Chapitre 4

État de l’art des données

Sommaire

4.1	Introduction	22
4.2	Conclusion	22

4.1 Introduction

Il arrive que l'état de l'art inclue aussi un chapitre sur les applications du même ordre, sur les modèles de données.

Ce n'est pas obligatoire d'avoir deux chapitres dans l'état de l'art ; en règle générale un seul chapitre suffit, parce que les données sont présentées dans la problématique.

4.2 Conclusion

Troisième partie

Système réalisé

Sommaire

5	Implémentation du système	27
5.1	Introduction	28
5.2	Conclusion	28
6	Expérimentations et résultats	29
6.1	Introduction	30
6.2	Conclusion	30
	Conclusion	31

Chapitre 5

Implémentation du système

Sommaire

5.1 Introduction	28
5.2 Conclusion	28

5.1 Introduction

Ici on présente le système qu'on a réalisé ; c'est une sorte de documentation technique, mais elle doit mettre en valeur les choix faits par l'étudiant (par exemple, si vous avez tenté quelque chose et que ça n'a pas marché, et finalement vous avez fait autrement, vous devez présenter vos tentatives infructueuses, elles témoigneront de la qualité de votre expérience).

5.2 Conclusion

Chapitre 6

Expérimentations et résultats

Sommaire

6.1 Introduction	30
6.2 Conclusion	30

6.1 Introduction

Ici on montre comment fonctionne le système réalisé ; une sorte de documentation utilisateur, mais qui montre les différents tests faits et les résultats de ces tests.

Idéalement, il devrait y avoir ici un benchmark, qui montre les progrès réalisés par rapport à d'autres solutions.

6.2 Conclusion

Conclusion

Ici il faut surtout attirer l'attention sur les progrès à faire, sur l'évolution du travail et les perspectives d'avenir.

Table des figures

Liste des tableaux

Table des matières

Introduction	5
Problématique	7
1 Le contexte de résolution du problème	11
1.1 Introduction	12
1.2 Conclusion	12
2 Le problème à résoudre	13
2.1 Introduction	14
2.2 Conclusion	14
État de l’art	15
3 État de l’art des techniques	19
3.1 Introduction	20
3.2 Indices externes	20
3.2.1 Taux global d’erreur de classification (T_{gec}) -sans coût- . . .	20
3.3 Conclusion	20
4 État de l’art des données	21
4.1 Introduction	22
4.2 Conclusion	22

Système réalisé	23
5 Implémentation du système	27
5.1 Introduction	28
5.2 Conclusion	28
6 Expérimentations et résultats	29
6.1 Introduction	30
6.2 Conclusion	30
Conclusion	31