

Rapport de Stage de Master 2

Signaux, Système, Image et Robotique

Juin 2007

Reconnaissance des formes et suivi de point
de fonctionnement. Application au diagnostic
d'aiguillage.

Encadrants :

M. Patrice AKNIN : directeur de recherche à l'INRETS

M. Samé ALLOU : chargé de recherche à l'INRETS

Présenté par :

Faïcel CHAMROUKHI

Table des matières

Remerciements	3
Présentation de l'INRETS	4
Chapitre 1. Introduction Générale	5
1. Contexte	5
2. Les appareils de voies	5
2.1 Définitions	5
2.2 Caractéristiques	5
2.3 Système d'aiguillage à manœuvre électrique	6
2.4 Télésurveillance des aiguilles	7
3. La manœuvre d'aiguille	8
3.1 Phase d'appel	9
3.2 Décalage du mécanisme	9
3.3 Déverrouillage de l'aiguille	9
3.4 Phase de translation	10
3.5 Phase de verrouillage	10
3.6 Phase de calage	11
3.7 Phase de friction	11
Conclusion	12
Chapitre 2. Présentation de la base de signaux utilisée	13
Introduction	13
1. Etiquetage des signaux	14
1.1 Classe 1 : Manœuvre sans défaut	14
1.2 Classe 2 : Manœuvre avec défaut tolérable	14
1.3 Classe 3 : Manœuvre avec défaut grave	15
2. Variabilité des signatures de défauts	15
2.1 Variabilité des signaux de la première classe	16
2.2 Variabilité des signaux de la deuxième classe	16
2.3 Variabilité des signaux de la classe 3	18
Conclusion	20
Chapitre 3. Segmentation et paramétrisation des signaux	21
Introduction à la Reconnaissance des Formes (RdF)	21
1. Méthodes de débruitage des signaux	21
1.1 Filtrage de Kalman	22
1.2 Filtrage par régression polynomiale avec fenêtre glissante	23
2. Segmentation	24
2.1 Détection du temps de début de manœuvre utile t_0	24
2.2 Définition des bornes des segments	25
3. Paramétrisation	26
3.1 Choix des paramètres	26

3.2	Rappel de la méthode des moindres carrés ordinaires.....	27
3.3	Paramètres	28
3.4	Diagramme de corrélation des variables	28
	Conclusion.....	29
Chapitre 4. Classification des signaux		30
	Introduction	30
1.	La classification supervisée.....	30
1.1	Définition	30
1.2	Normalisation des paramètres	31
2.	Mise en œuvre des différentes méthodes de classification supervisée.....	31
2.1	Classification directe	31
2.1.1	Classification par k-plus proches voisins	31
2.1.2	Classifieur Nearest Mean (la moyenne la plus proche).....	34
2.2	Approches probabilistes	37
2.2.1	Rappels sur la classification ou discrimination Bayésienne.....	37
2.2.2	Discrimination par mélange de densités gaussiennes	41
2.3	Classification par réseau de neurones	42
2.3.1	Définition d'un réseau de neurones.....	42
2.3.2	Perceptron Multicouches (MLP).....	44
2.3.3	Apprentissage	45
2.3.4	Résultats	47
2.3.5	Interprétations.....	49
	Conclusion.....	49
Conclusion Générale		51
Bibliographie.....		52

Remerciements

Je tiens à remercier M. Patrice Aknin, directeur de recherche à l'Institut National de Recherche sur les Transports et leur Sécurité (INRETS), pour m'avoir permis d'effectuer ce stage et m'avoir encadré de façon très efficace. Ses remarques et ses conseils ont été les fils conducteurs de cette étude.

Je remercie également M. Allou Samé, chargé de recherche INRETS, pour la multitude de conseils qu'il m'a donnés. Je le remercie également pour son soutien qui a fait avancer cette étude sur de « bons rails ».

J'exprime surtout mes remerciements à la SNCF, en particulier à M. Marc Antoni, qui a su me transmettre une partie du savoir faire de la SNCF, fort utile dans cette étude.

Présentation de l'INRETS

L'Institut National de Recherche sur les Transports et leur Sécurité (INRETS) est un Etablissement Public à caractère Scientifique et Technologique (EPST), créé par décret interministériel du 18 septembre 1985 et placé sous la tutelle des ministères chargés de la recherche et des transports. L'INRETS compte plusieurs laboratoires dont le Laboratoire des Technologies Nouvelles (LTN) où ce stage a eu lieu.

Les compétences du LTN concernent d'une part les aspects électromécaniques des transports guidés (sustentation, guidage, propulsion) avec une approche technologique sur les composants et les méthodes innovantes de diagnostic, et d'autre part une approche globale des systèmes de transport, plus particulièrement des systèmes guidés. Ces compétences sont regroupées en quatre pôles de recherche :

- Transport et Transformation de l'Energie pour la Traction Electrique ;
- Dynamique des Véhicules Guidés ;
- Diagnostic et Maintenance des Systèmes de Transports Guidés ;
- Systèmes de Transport Guidé Urbains et Interurbains ;

Le pôle "Diagnostic et Maintenance des Systèmes de Transports Guidés" du LTN a vu le jour en 1997. Depuis ces dernières années, ce pôle intensifie ses activités sur les thèmes du diagnostic et de la maintenance, en s'appuyant sur les disciplines académiques du traitement de signal, de la reconnaissance des formes, de la statistique, du diagnostic et de la fiabilité [INT].

Le stage s'est effectué dans le cadre d'un partenariat entre le pôle « Diagnostic et maintenance des systèmes de transports guidés » et la direction de l'infrastructure de la SNCF.

Chapitre 1. Introduction Générale

1. Contexte

La télésurveillance des organes de l'infrastructure ferroviaire est un sujet d'actualité pour les opérateurs notamment la SNCF. La question est comment éviter l'arrêt d'une installation complexe d'exploitation quasi permanente provoqué par le dysfonctionnement d'un composant.

La maintenance prévisionnelle semble être la seule solution valable qui permette de prévenir de tels incidents. En effet, avec cette approche, la défaillance est instantanément détectée et communiquée au service concerné, ce qui permet, soit d'intervenir immédiatement, soit de programmer des opérations de remplacement du dispositif qui présente un risque de panne. La sécurité et la qualité de service s'en trouvent donc améliorées si des décisions de maintenance préventive sont prises judicieusement, suite aux alarmes générées et envoyées automatiquement par les postes de contrôle-commande des trains. La génération de ces alarmes nécessite un travail conséquent en amont, afin de distinguer sur les signaux mesurés, l'apparition de défaillances et d'en suivre les évolutions.

Ce sujet de stage concerne le diagnostic des moteurs d'aiguillage de voies par une approche « Reconnaissances des Formes » afin d'améliorer le dispositif de maintenance préventive des systèmes d'aiguillage de la SNCF. A partir d'une base de signaux mesurés dans des conditions réelles, l'objectif de ce travail est de trouver un espace de représentation des signaux et de mettre en œuvre différentes méthodes de classification supervisée pour apprendre les états de fonctionnement des aiguillages.

2. Les appareils de voies

2.1 Définitions

Un appareil de voie est un dispositif permettant d'assurer la continuité de la voie pour un itinéraire choisi parmi divers itinéraires divergents ou sécants. Les aiguillages sont des exemples d'appareils de voie. Ils permettent de guider les trains dans différentes directions à l'aide d'aiguilles. Les deux voies vers lesquelles le train peut se diriger sont sécantes entre elles au niveau de ce qu'on appelle le « cœur d'aiguille ». C'est à ce niveau qu'il existe une discontinuité dans le chemin de roulement de la roue du train, afin de permettre le passage du « boudin » qui assure le guidage de la roue sur le rail.

2.2 Caractéristiques

Les appareils de voie comprennent :

- les branchements à 2 ou 3 voies (à déviation à droite, à déviation à gauche ou symétriques) permettant la ramification d'une voie,
- les traversées permettant le cisaillement d'une voie par une autre,
- les traversées jonctions permettant le cisaillement et la ramification des voies.

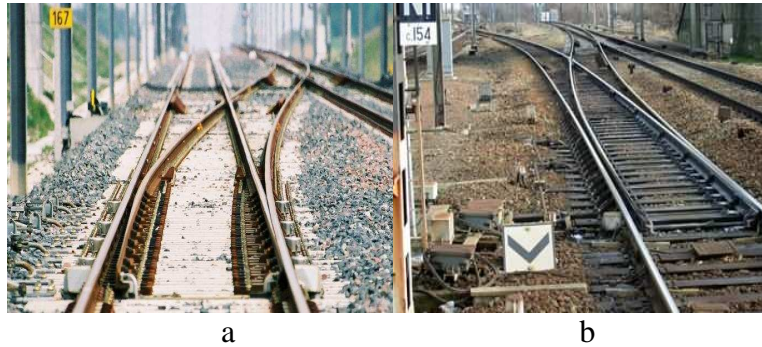


Figure 1 : a. Jonction simples, b. Aiguille commandée à distance.

2.3 Système d'aiguillage à manœuvre électrique

L'aiguillage à manœuvre électrique est composé :

- d'un mécanisme électrique comportant essentiellement :
 - un moteur électrique,
 - un limiteur de couple à friction,
 - un système d'entraînement qui assure le calage,
 - un bras de manœuvre pour translation des aiguilles,
 - un dispositif de manœuvre de secours.
- de deux demi-aiguillages constitués essentiellement chacun par :
 - un contre-aiguille,
 - des coussinets de glissement solidaires du contre-aiguille et fixés sur des traverses.
- d'une tringle d'écartement et éventuellement d'une attaque multiple.

Lors de la manœuvre électrique d'un appareil de voie, l'énergie nécessaire est fournie par une alimentation électrique (380 V en alternatif ou 150 V en continu). Ainsi, la mesure et l'analyse de la variation temporelle de cette énergie consommée au cours d'une manœuvre permettent de détecter les principales causes de dysfonctionnement avant qu'un raté de commande ne survienne réellement. La figure 2 montre un moteur d'aiguille à manœuvre électrique [DOC].

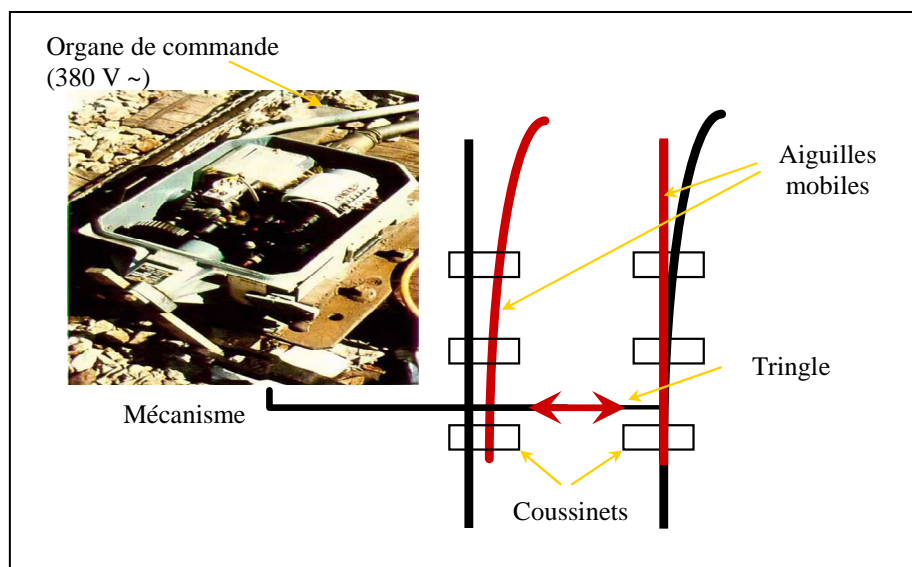


Figure 2 : Aiguillage à manœuvre électrique

2.4 Télésurveillance des aiguilles

Ce système présente essentiellement trois parties :

1. Une partie capteurs : pour l'acquisition, par des capteurs locaux, des grandeurs physiques représentant l'état de fonctionnement des différentes composantes sur les voies notamment les moteurs d'aiguilles et l'aiguillage.
2. Une partie réseau pour permettre l'accès à distance aux signaux mesurés.
3. Une troisième composante pour le contrôle commande qui consiste en la gestion des alarmes, intervention en cas d'incident.

L'information utile est représentée sous forme d'un graphique de la puissance consommée en fonction du temps, caractérisant la dernière manœuvre effectuée. L'acquisition des signaux des manœuvres se fait par des capteurs embarqués sur les appareils de voie. La gestion de ces signaux se fait à partir des postes de contrôle-commande des trains. Plusieurs modes d'acquisition des signaux peuvent être envisagés :

- l'acquisition par un oscilloscope portable à mémoire permettant de visualiser le courant appelé par le mécanisme lors d'une manœuvre d'aiguille. Le dispositif est limité à la présence d'un agent de maintenance sur le site.
- l'acquisition par un équipement dédié, indépendant de la télésurveillance. L'installation est temporaire ou permanente, en fonction de l'importance de l'aiguille, de la période de l'année ou des défauts précédemment constatés. Les informations sont disponibles sur le site ou à distance, par téléinterrogation.
- L'acquisition par le biais d'équipements exploités par un système de télésurveillance. Les informations sont disponibles à distance depuis un poste dit « poste V non dédié ».

La figure 3 montre le système de télésurveillance de La SNCF avec ses trois principales fonctionnalités.

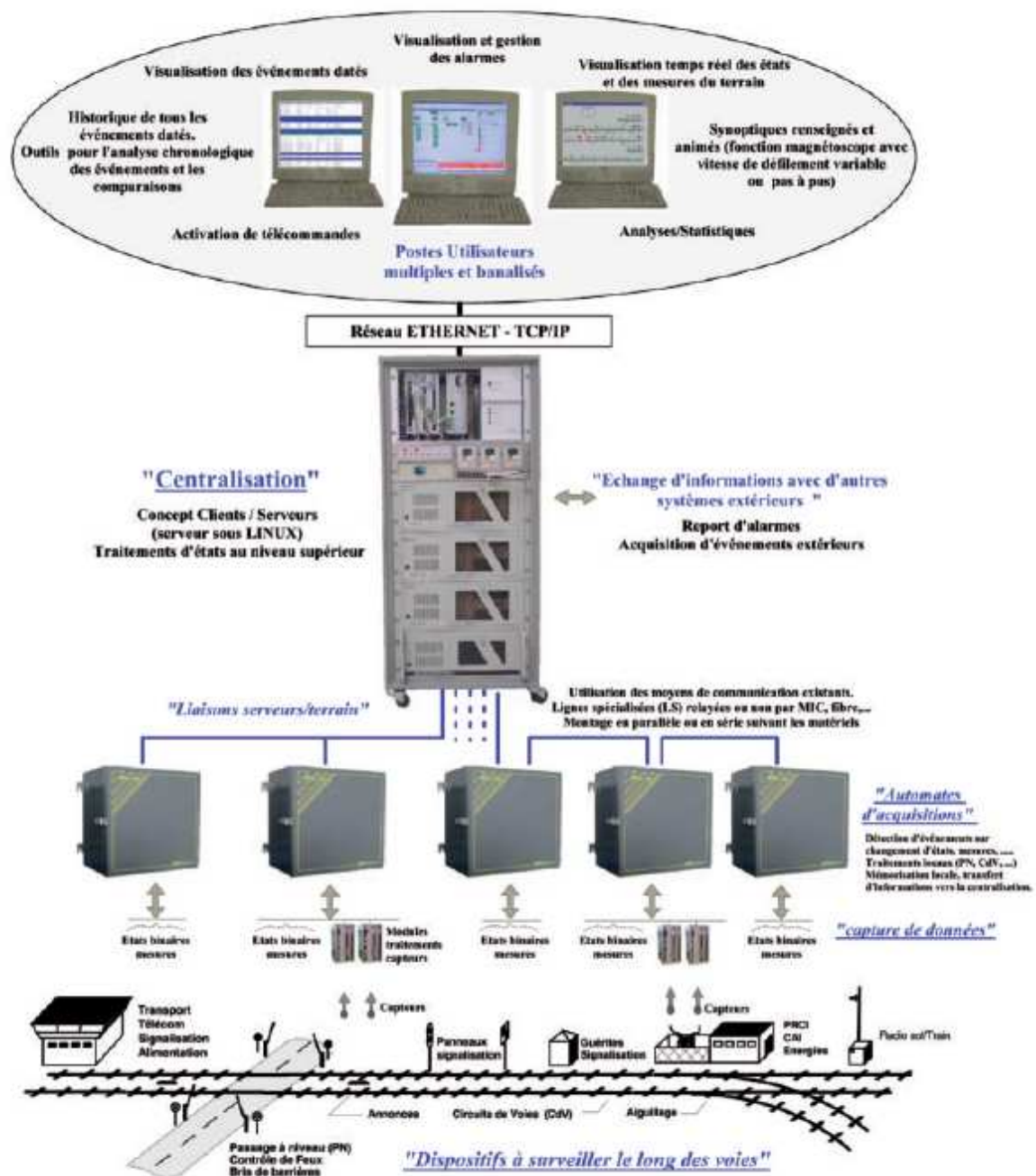


Figure 3 : Système de télésurveillance de l'infrastructure de la SNCF.

3. La manœuvre d'aiguille

Une manœuvre d'aiguille est constituée de mouvements mécaniques des différents organes liés à l'aiguille, qui sont mobilisés successivement. Ces mouvements se traduisent, sur le signal de puissance consommée en fonction du temps, par différentes phases de fonctionnement

3.1 Phase d'appel

La pointe (effort élevé au début de signal) caractérise les pertes dans le câble d'alimentation et la consommation du moteur bloqué (voir figure 4). La durée de cette phase et la puissance consommée dépendent de la longueur et du type du câble (résistance et inductance équivalente du câble) la durée peut varier typiquement de 50 ms à 300 ms [DOC].

3.2 Décalage du mécanisme

Dès la fin de la pointe initiale, le moteur déplace le doigt de calage sur la portée de calage, de manière à préparer la manœuvre utile du mécanisme. Dans la première moitié de cette phase, le moteur s'oppose au frein de calage. Dans la seconde, le moteur entraîne le doigt de calage se déplaçant librement sur la portée de calage. La valeur de l'effort relevé sur cette moitié peut servir de référence (moteur tournant à vide). En effet, ce niveau est indépendant de la charge du mécanisme. Cette phase dure environ 300 ms [DOC]. La figure 4 décrit la puissance consommée durant la phase de décalage.

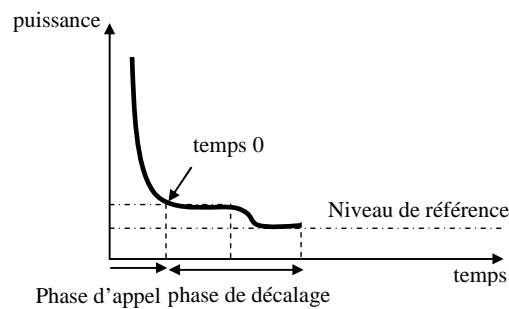


Figure 4 : courbe typique d'un décalage

3.3 Déverrouillage de l'aiguille

Cette phase correspond à l'effort de manœuvre nécessaire pour déverrouiller la lame d'aiguille fermée et décaler la lame ouverte jusqu'à une position autorisant la translation des lames d'aiguille. La consommation dans cette phase correspond en plus de la consommation à vide, à l'effort de rotation des C (dans le cas de l'existence de VCC : Verrou Carter Coussinet) et, si l'attaque multiple est raccordée à la tringle des VCC, à l'effort de translation du talon des lames de l'appareil de voie.

Une sur consommation pour cette zone de fonctionnement peut traduire, selon sa forme, par exemple, un défaut de graissage, une action de traction de la lame appliquant la tête de C sur la pièce de bronze. La figure 5 montre la courbe de la puissance consommée lors de cette phase [DOC]. Cette phase dure environ 830 ms.

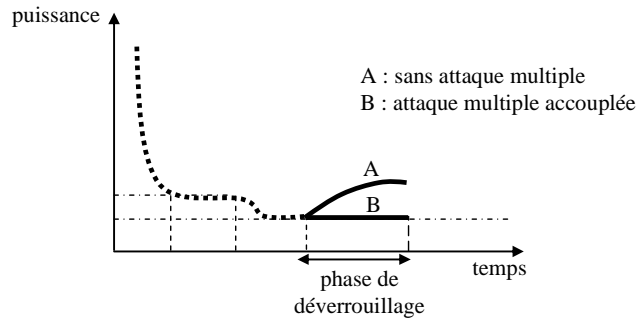


Figure 5 : courbe typique d'un déverrouillage

3.4 Phase de translation

Cette phase, présentée dans la figure 6, correspond à l'entraînement simultané des deux lames d'aiguille. Dans le cas d'une liaison de l'attaque multiple à la tringle d'accouplement, il est à noter que la manœuvre de la pointe des lames est exécutée plus rapidement que le talon de l'aiguille. Un dérèglement de l'attaque multiple (par le froid par exemple) se répercute sur l'effort de manœuvre. Dans le cas contraire l'effort de manœuvre des deux lames est appliqué brutalement dès la fin de la zone de déverrouillage. Si un problème oscillatoire apparaît sur l'effort de manœuvre, il se traduit par un défaut de graissage du talon des lames, provoquant un déplacement en cascade de celles-ci. Si une pointe d'effort apparaît en fin de translation, elle traduit une sur course (défaut de réglage) de la tringle de commande de l'appareil ou un dérèglement d'une butée de l'appareil [DOC]. Cette phase dure environ 2.14 S.

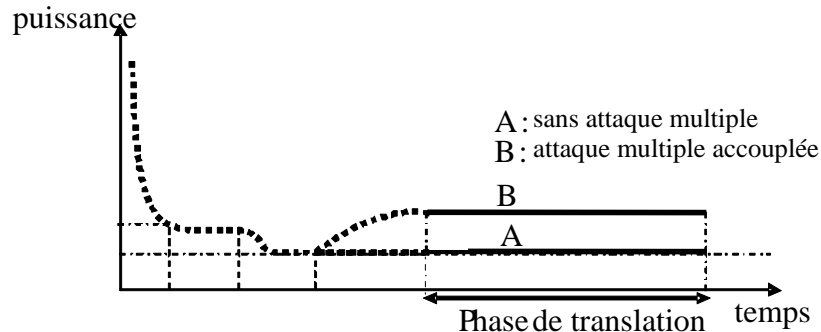


Figure 6 : courbe typique d'une translation

3.5 Phase de verrouillage

Cette phase correspond à l'effort de manœuvre nécessaire pour verrouiller la nouvelle lame fermée et décaler la lame ouverte. La consommation du mécanisme correspond normalement, en plus de la consommation à vide, à l'effort de rotation des C (dans le cas de l'existence des VCC) et, si l'attaque multiple est raccordée à la tringle d'accouplement des VCC de verrouillage, à l'effort de translation finale du talon des lames de l'appareil de voie. La puissance consommée lors de cette phase de fonctionnement est présentée dans la figure 7. Une surconsommation dans cette phase de fonctionnement peut traduire selon sa forme, un défaut de graissage, une action de traction de la lame traduisant une difficulté de verrouillage. Cette phase dure environ 830 ms [DOC].

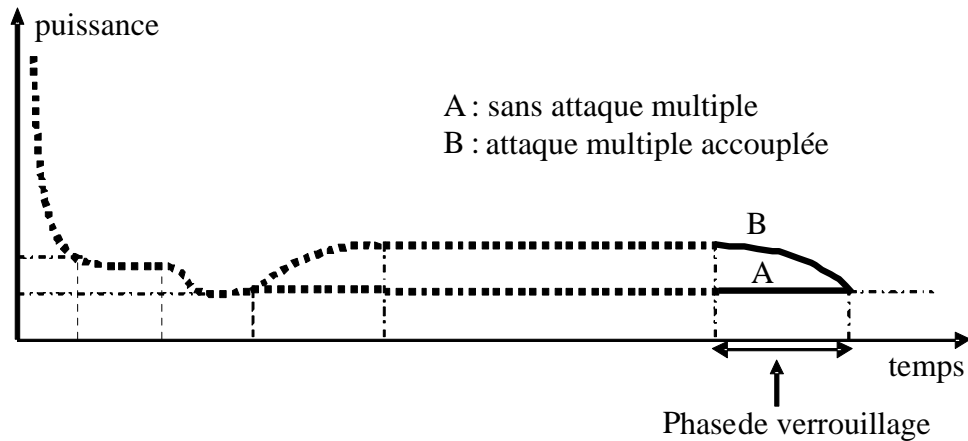


Figure 7: courbe typique d'un verrouillage

3.6 Phase de calage

De même que pour la phase de décalage, la phase de calage du mécanisme se traduit par une courbe typique de puissance consommée lors de cette phase de fonctionnement. Dans tous les cas, le niveau de l'effort de calage doit être proche du niveau de référence fixé par la phase de décalage [DOC]. Une courbe typique de la puissance active lors de cette phase est décrite par la figure 8. Cette phase dure environ 300 ms.

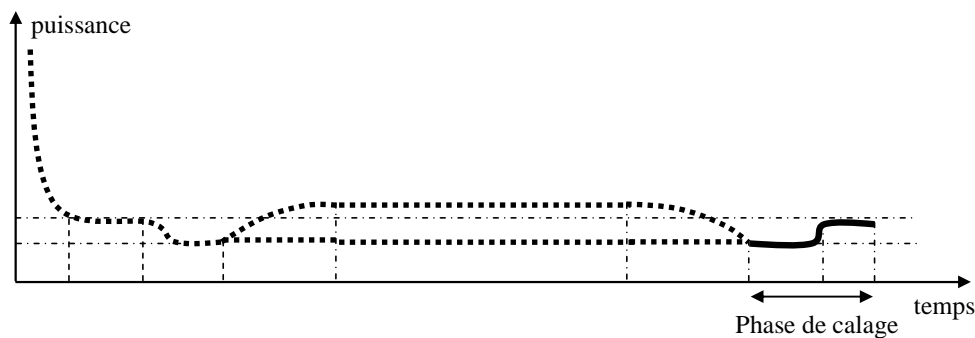


Figure 8 : courbe typique d'un calage

3.7 Phase de friction

Cette phase de fonctionnement traduit la mise en œuvre du dispositif de limitation de couple destiné à protéger le moteur et les engrenages du mécanisme en cas de blocage de la lame par un obstacle empêchant la fin de la manœuvre, la limitation est ainsi en œuvre. Le temps d'une manœuvre est de 4.25 s (sans la marge de sécurité) le temps d'activation de la limitation de couple [DOC]. La figure 9 montre la courbe de la puissance consommée lors de la phase de friction.

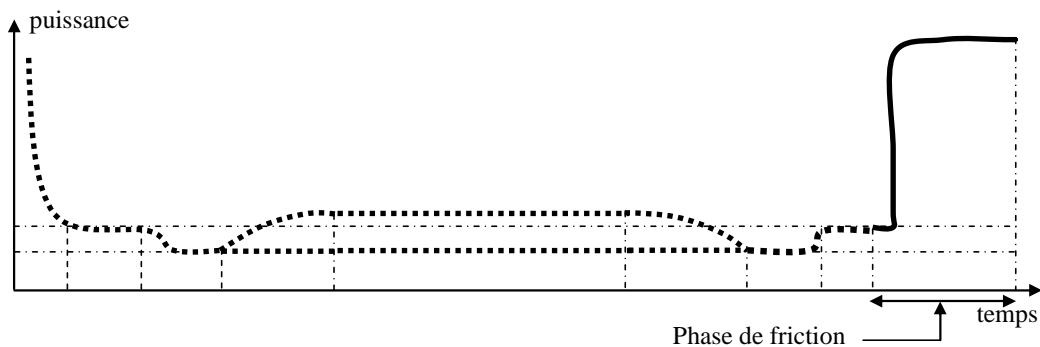


Figure 9 : courbe typique d'une Friction

Conclusion

On a décrit dans ce chapitre le principe de fonctionnement des différentes phases d'un mécanisme d'aiguillage à manœuvre électrique, à travers la caractéristique, puissance consommée en fonction du temps. Ces différentes phases de fonctionnement sont représentées sur la figure 10.

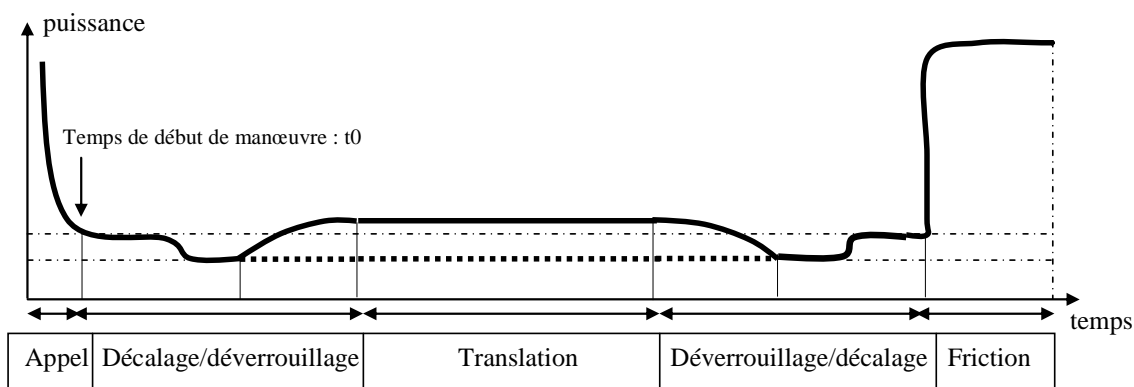


Figure 10 : Manœuvre entière

La compréhension du principe de fonctionnement des différentes phases de l'aiguillage est indispensable pour mettre en œuvre un système de diagnostic fiable et capable d'envisager différents types de défauts de l'appareil de voie.

Après avoir détaillé le fonctionnement du mécanisme, vient l'examen de la variabilité des différents types de défauts susceptibles de survenir sur une manœuvre d'aiguilles, l'objectif du chapitre suivant.

Chapitre 2. Présentation de la base de signaux utilisée

Introduction

La base de données dont on dispose est formée de 119 des signaux de manœuvres d'aiguillage fournis par la SNCF. Pour certaines manœuvres, on dispose aussi de signaux correspondant à la manœuvre retour (en sens inverse). Ces signaux ont été acquis sur un site du réseau ferré de la SNCF. Les mécanismes d'aiguillage de ce site présentent les caractéristiques suivantes :

- manœuvre à mécanisme électrique commandant une aiguille TGV,
- système d'aiguillage avec VCC (Verrou Carter Coussinet),
- mécanisme muni d'accouplement de l'attaque multiple.

Ces mesures ont été prises après avoir généré des défauts sur l'aiguillage, de façon à avoir une base de données qui représente le mieux possible la variabilité des signatures. Les signaux représentent la tension (voir figure 11), le courant (voir figure 12) et la puissance consommée en fonction du temps (voir figure 13).

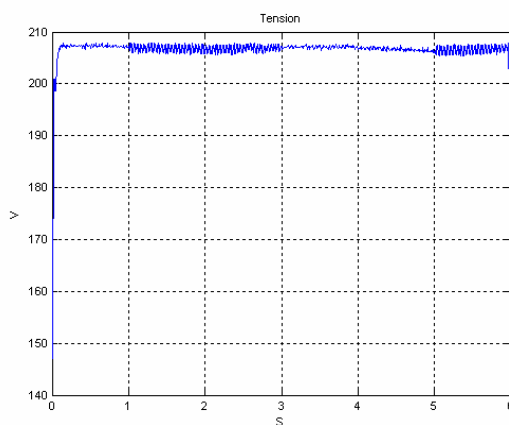


Figure 11 : Signal de tension appliquée en fonction de temps

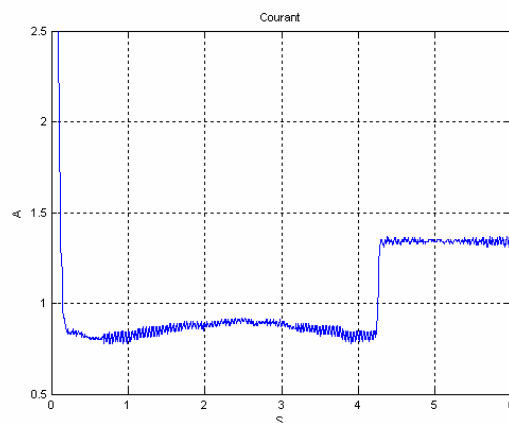


Figure 12 : Signal de courant absorbé en fonction de temps

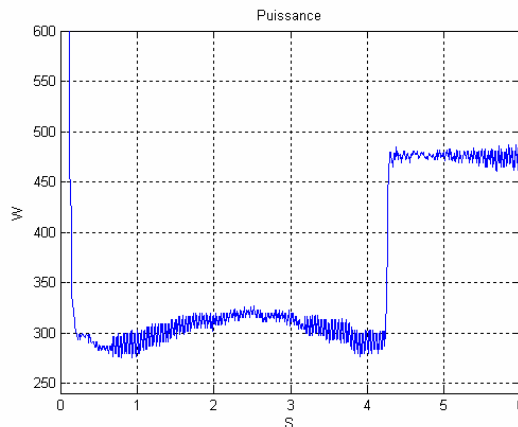


Figure 13 : Signal de puissance consommée en fonction de temps

Pour effectuer le diagnostic, seul le signal de puissance (ou de manière équivalente le courant) a été utilisé. Le courant a servi pour la détection du temps de début de la manœuvre utile.

1. Etiquetage des signaux

Une réunion à la SNCF le 26 Avril 2007 a permis d'étiqueter les 119 signaux mis à disposition.

1.1 Classe 1 : Manœuvre sans défaut

Cette classe regroupe les signaux correspondant à des manœuvres « normales » et ne présentant pas de défauts, qui peuvent être critiques pour l'aiguillage des voies. 29 signaux de cette classe sont disponibles. Un exemple de cette classe est présenté dans la figure 14

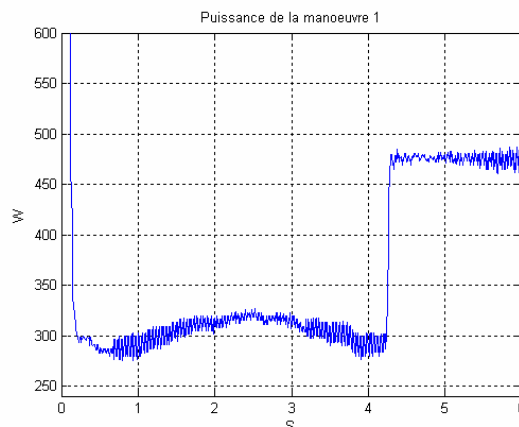


Figure 14 : Manœuvre sans défaut

1.2 Classe 2 : Manœuvre avec défaut tolérable

Cette classe regroupe les manœuvres qui présentent des défauts tolérables (26 manœuvres). Ces défauts ne nécessitent pas d'action de maintenance immédiate mais leur signalement

permet d'envisager, à moyen terme, un entretien de l'appareil de voie. La figure 15 montre un exemple de cette classe.

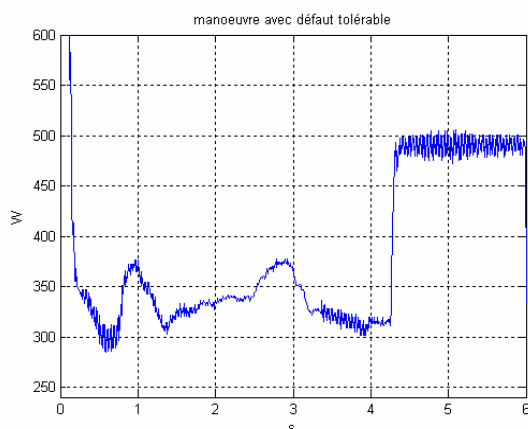


Figure 15 : Manœuvre avec défaut tolérable

1.3 Classe 3 : Manœuvre avec défaut grave

Cette classe regroupe l'ensemble des manœuvres présentant un défaut grave, nécessitant une intervention immédiate. Nous disposons de 64 manœuvres de cette classe, comme le cas présenté par la figure 16. Généralement, un signal d'alarme est déclenché pour les manœuvres de cette classe.

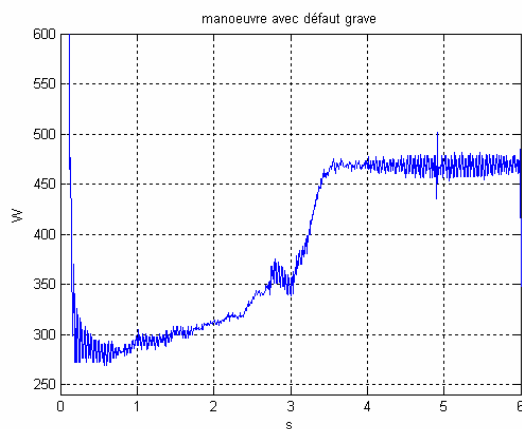


Figure 16 : Manœuvre avec défaut grave

2. Variabilité des signatures de défauts

La variabilité des signatures, pour les classes 2 et 3 est très importante et celles-ci peuvent s'écarter sensiblement de celles représentées sur les figures 15 et 16. Cette partie donne quelques exemples de signatures différentes pour les signaux des classes 2 et 3. L'interprétation de ces signatures a été effectuée par des agents experts de la SNCF. Ci-dessous les signaux avec leurs classes et des informations sur les types de défauts :

Les défauts qui peuvent survenir lors d'un aiguillage sont majoritairement répartis en deux catégories :

- Les défauts de type mécanique comme le défaut de graissage, mauvais réglage de l'attaque multiple, ou les défauts liés au froid,
- Les défauts électriques tels que l'échauffement du limiteur de couple.

Les figures qui suivent représentent la variabilité des différentes classes et la variabilité des signatures de défauts dans chaque classe.

2.1 Variabilité des signaux de la première classe

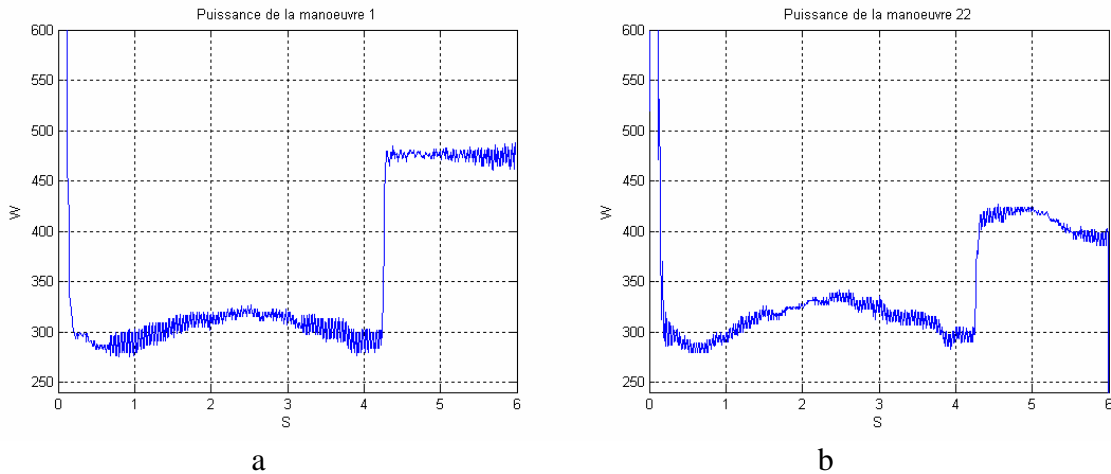


Figure 17: Manœuvres de la classe 1

La figure (17.a) représente la courbe de puissance type pour une manœuvre normale. La puissance consommée lors de chaque phase de fonctionnement ainsi que le temps que dure chaque phase sont proches des valeurs optimales. Le niveau de l'effort dans la phase de verrouillage, translation et déverrouillage un peu élevé par rapport à la référence (moteur à vide) est dû à l'accouplement de l'attaque multiple.

La seconde manœuvre appartenant à la classe 1 (figure17.b) présente un niveau d'effort légèrement plus élevé que la manœuvre (a). Cette différence très légère n'engendre aucun changement pour la décision sur la classe de la manœuvre. Pour cette manœuvre on remarque aussi la présence d'une légère décroissance en fin de limitation de couple à cause d'un échauffement du limiteur de couple.

2.2 Variabilité des signaux de la deuxième classe

La manœuvre de la figure 18.a présente un défaut tolérable, l'effort relativement élevé de la phase de translation qui entraîne une marge insuffisante entre le niveau d'effort en limitation de couple et celui de la phase de translation. Une intervention non nécessairement immédiate est à prévoir.

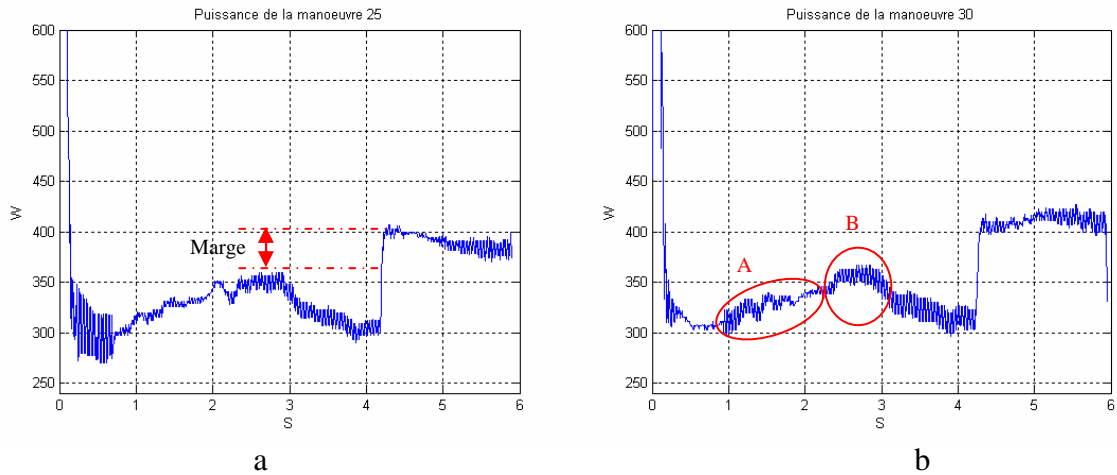


Figure 18 : Manœuvres de la classe 2

Sur le signal de la figure 18.b, les oscillations en phase de translation traduisent un léger mauvais graissage du talon de l'aiguille. Ce phénomène est appelé « broutage de l'aiguille ». Le léger sur-effort indique également un manque de graisse ou un dérèglement de l'attaque multiple.

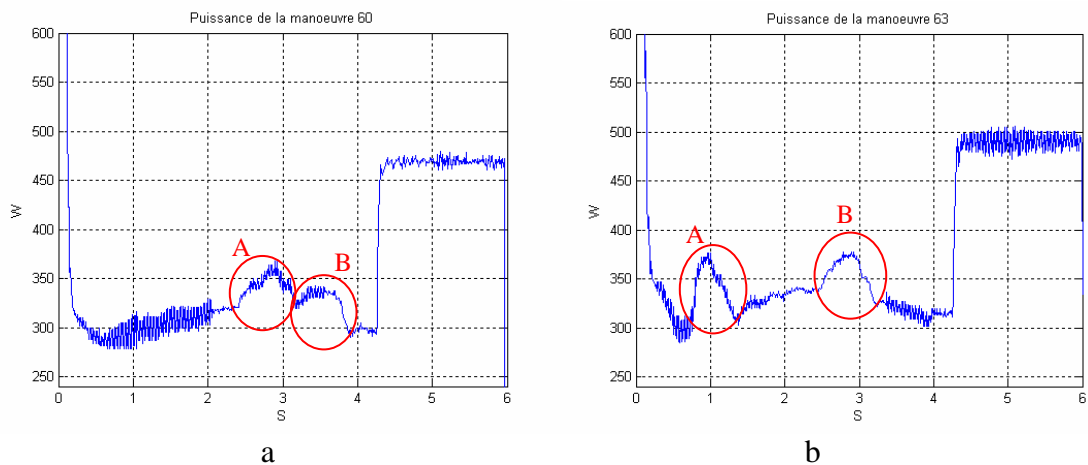


Figure 19 : Manœuvres de la classe 2

Pour la manœuvre de la figure 19.a, on remarque :

- A : un sur effort dû à une sur course de l'attaque multiple,
- B : un sur effort durant la course de verrouillage VCC.

Pour la manœuvre représentée dans la figure 19.b, on peut observer :

- A : un sur effort en phase de déverrouillage : Problème graissage ou grippage VCC,
- B : un sur-effort en fin de translation dû à une sur-course de l'attaque multiple.

2.3 Variabilité des signaux de la classe 3

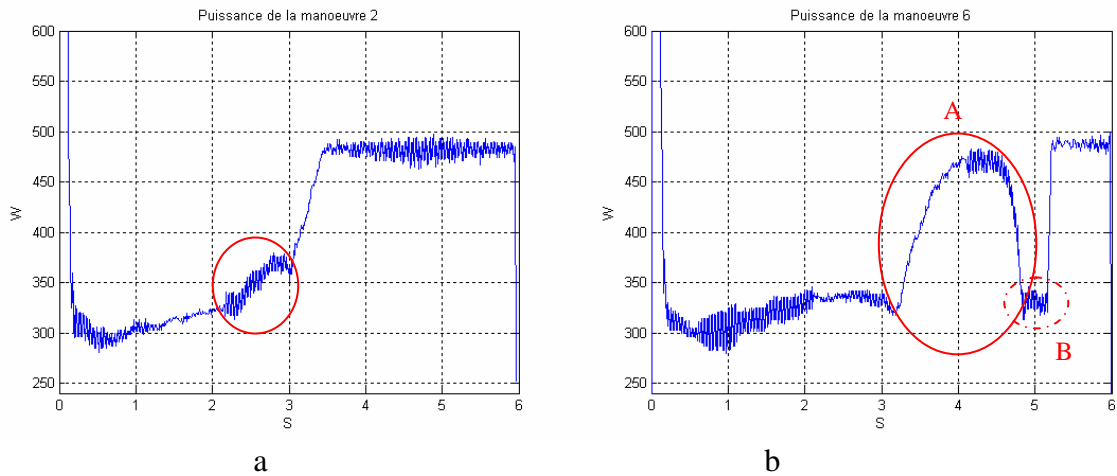


Figure 20 : Manœuvres de la Classe 3

La manœuvre présentée dans la figure 20.a présente un problème en fin de translation et à la phase de verrouillage/calage, qui n'a pas eu lieu. La phase de translation se prolonge jusqu'à la mise en œuvre de la limitation de couple.

Pour la manœuvre de la figure 20.b on site les défauts suivants :

- A : un obstacle entre les deux paulvés et le patinage de la friction sous l'effort (limite de dérochement),
- B : l'aiguille est verrouillée est calée ce qui signifie l'aboutissement de la manœuvre, mais on encourt un grand risque pour la manœuvre qui suivra.

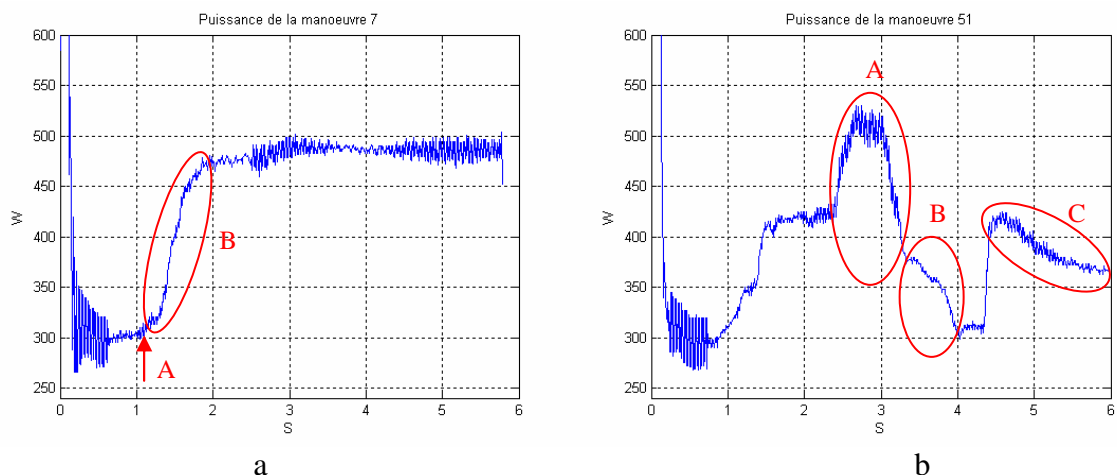


Figure 21 : Manœuvres de la classe 3

La manœuvre de la figure 21.a présente les défaillances suivantes :

- A : la rupture de la tringle de manœuvre qui assure la translation des lames de l'aiguille et la phase de translation n'a pas eu lieu,
- B : le patinage de la friction sous l'effort suite à la rupture de la tringle.

Pour la manœuvre de la figure 21.b, on observe :

- A : un sur effort traduisant un grave problème au niveau de l'attaque multiple,
- B : un effort anormal en phase de verrouillage dû à un mauvais graissage ou à un frottement entre tête du C et pièce de bronze,
- C : le limiteur de couple perd sa pertinence (échauffement du limiteur de couple mécanique classique).

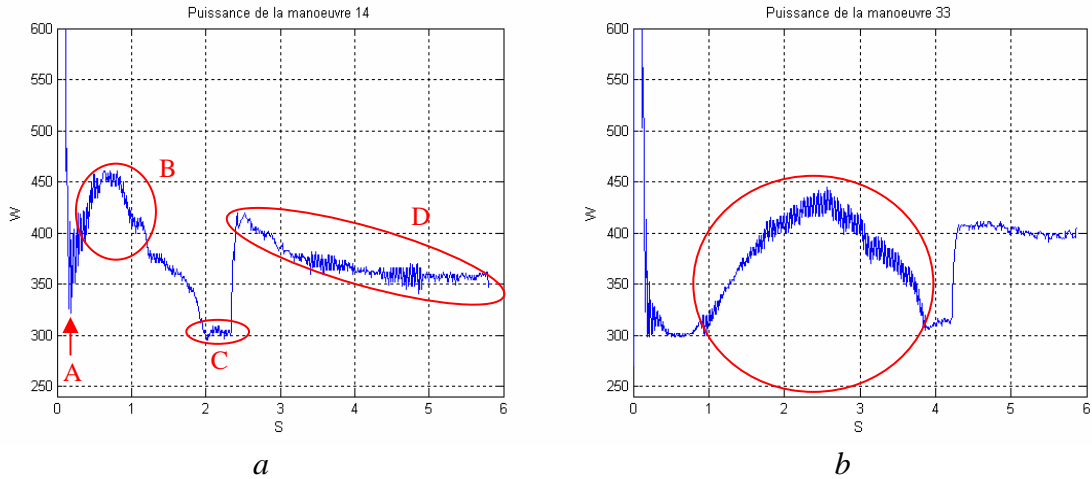


Figure 22 : manœuvres de la classe 3

La manœuvre représentée dans la figure (22.a) est une manœuvre retour d'une manœuvre aller avec défaut grave. Pour cette manœuvre, on observe que :

- A : le début de cette manœuvre est marqué par l'absence du décalage/déverrouillage,
- B : le sur effort très élevé lors de la phase de translation,
- C : l'aiguille est verrouillée et calée ce qui traduit que la manœuvre a abouti mais le risque pour le prochain aiguillage est élevé,
- D : le limiteur de couple perd son efficacité d'après la décroissance de la puissance consommée lors de la phase de friction. (échauffement du limiteur de couple mécanique classique).

La manœuvre présentée dans la figure (22.b) présente un sur effort élevé en phase de translation et de verrouillage, la forme est typique d'un défaut de graissage.

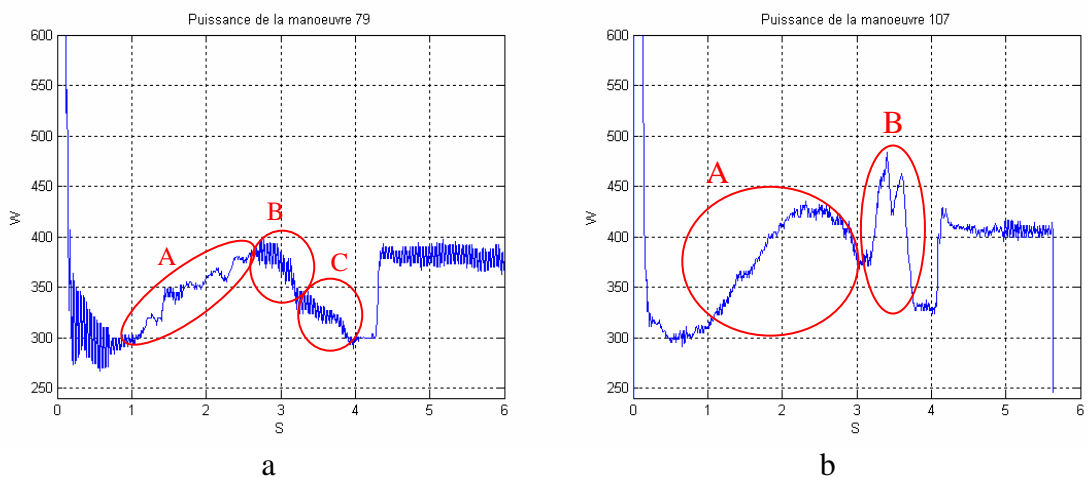


Figure 23 : Manœuvres de la Classe 3

Pour le signal de manœuvre représenté dans la figure (23.a) on remarque la présence des défauts suivants :

- A : le broutage de l'aiguille traduit un mauvais graissage du talon de l'aiguille,
- B : un léger problème au niveau de l'attaque multiple,
- C : sur effort en phase de verrouillage mauvais graissage ou frottement entre tête du C et pièce de bronze.

La manœuvre représentée dans la figure (23.b) présente :

- A : un sur effort lors de la translation qui est dû à un mauvais graissage,
- B : un pic d'effort en début de la phase de verrouillage dû à la présence d'un obstacle au talon de lame d'aiguille (probablement un caillou qui s'est écrasé).

Conclusion

On a détaillé dans ce chapitre la variabilité des signatures des différents défauts, ce qui représente une collecte de connaissances à priori qui nous servira à faire le bon choix des caractéristiques lors de la paramétrisation des signaux.

Vu la variabilité des phases de fonctionnement qui interviennent lors d'une manœuvre d'aiguille, une prise en compte des principales caractéristiques de chaque phase au niveau de la paramétrisation des signaux de manœuvre, est l'objectif du chapitre suivant.

Chapitre 3. Segmentation et paramétrisation des signaux

Introduction à la Reconnaissance des Formes (RdF)

La reconnaissance des formes consiste à identifier ou classer des formes ou objets en se basant sur certaines de leurs caractéristiques. Cette approche est maintenant utilisée dans plusieurs domaines d'activités. En diagnostic ferroviaire, la RdF a été utilisée pour diagnostiquer le circuit de voie [DEBIOLLES 07], [OUKHELLOU 97]. Les différents modules d'un système de RdF sont représentés par la figure 24.

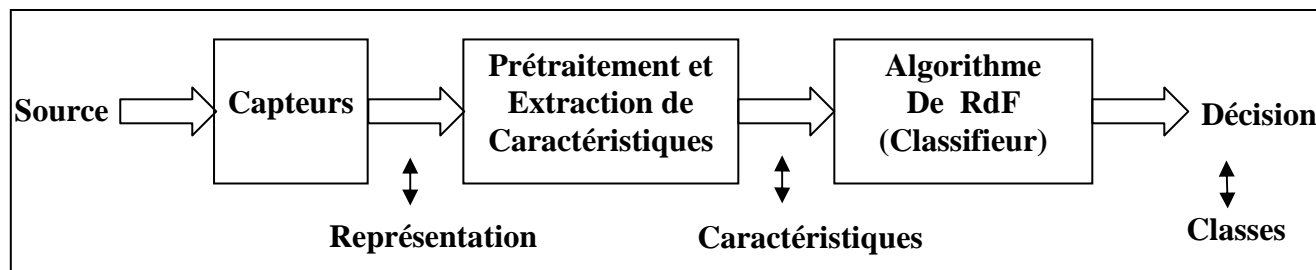


Figure 24: Schéma d'un système de Reconnaissance de Formes

Le premier module d'un système de RdF est celui de l'acquisition de données brutes, sans aucun prétraitement. L'espace de ces observations est souvent de dimension élevée. La réduction de la dimension se fait par extraction de vecteur forme pour chaque observation. Ce vecteur représente les caractéristiques les plus représentatives des observations. Cette dernière étape est aussi appelée l'étape de paramétrisation.

La classification est l'étape qui consiste à associer à chaque signal traité représenté dans l'espace de caractéristiques, une classe de l'espace de décision contenant les différentes classes d'appartenance des données.

Dans le cadre de notre application de diagnostic de moteur d'aiguillage, la paramétrisation des signaux de mesure passe par un débruitage et une segmentation préalables des signaux, suivant les différentes phases de manœuvre décrites dans le premier chapitre. Nous détaillons dans la suite les étapes de débruitage, de segmentation et finalement la paramétrisation des signaux.

1. Méthodes de débruitage des signaux

Pour avoir une bonne segmentation, l'élimination du bruit superposé au signal est indispensable. Pour cela, plusieurs méthodes de débruitage ont été envisagées.

1.1 Filtrage de Kalman

Un article sur le diagnostic des moteurs d'aiguilles [PEDREGAL 04], traite du problème de détection de défaut par observation de l'évolution du coefficient de corrélation entre le nouveau signal de manœuvre acquis (à tester) et un signal de manœuvre modèle. Avant d'effectuer le diagnostic, cet article propose une méthode de débruitage de signal par le filtre de Kalman. Nous avons implémenté cette méthode de filtrage et nous avons obtenu les résultats suivants présentés dans la figure 25 et les tableaux 1 et 2.

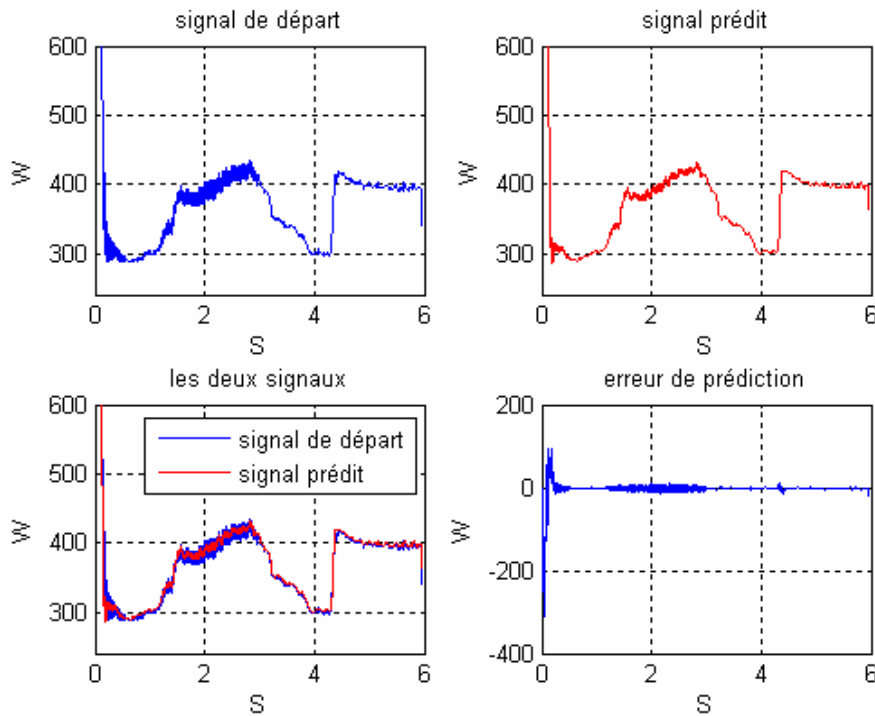


Figure 25 : Résultats du filtrage de Kalman

Le signal prédit représente le signal filtré, l'erreur de prédiction représente le bruit.

On utilise les notations suivantes :

- Erreur : (bruit) $\varepsilon = x - \hat{x}$ où x le signal observé, et \hat{x} le signal débruité ;
- Moyenne de l'erreur : $\bar{\varepsilon} = \frac{1}{N} \sum_{i=1}^N \varepsilon_i$;
- Ecart type : $\sigma = \frac{1}{N} \sum_{i=1}^N (\varepsilon_i - \bar{\varepsilon})^2$;
- La puissance de l'erreur : $P_\varepsilon = 10 \log_{10} \left(\sum_{i=1}^N \varepsilon_i^2 \right)$.

En s'appuyant sur ces notations, les résultats obtenus pour le signal présenté dans la figure 27 sont présentés dans le tableau 1.

$\bar{\varepsilon}$	σ	P_ε
-2.91	21.01	54.27

Tableau 1 : Résultat du filtrage pour un signal

Les résultats moyennés sur tous les signaux sont présentés dans la table 2 :

$\bar{\varepsilon}$	σ	P_{ε}
-8.64	89.94	65.21

Tableau 2 : Résultat du filtrage sur tous les signaux

On peut remarquer que le signal prédit, qui est censé fournir le bruit le moins for possible, présente un bruit de variance moins forte que celui du signal d'origine. Cependant ce bruit reste élevé, en effet le lissage n'est pas efficace. Cela peut être attribué au fait que le modèle proposé est mieux adapté à des signaux moins bruités que ceux de notre application. De plus, la méthode que l'article propose est conçue pour effectuer un diagnostic selon l'évolution du coefficient de corrélation entre la manœuvre observée et une manœuvre de référence. Dans notre cas, nous ne disposons pas de courbe de référence. Donc l'implémentation de cet article était pour faire le filtrage des signaux et d'en évaluer les résultats.

1.2 Filtrage par régression polynomiale avec fenêtre glissante

La méthode de débruitage par régression polynomiale avec fenêtre glissante consiste à estimer l'échantillon du signal x_i à l'instant t_i de la manière suivante :

On choisit une fenêtre de taille M centrée sur t_i ; on effectue ensuite une régression polynomiale (de degré 1 ou 2) par la méthode des Moindres Carrés. Finalement, la valeur x_i est la moyenne des points dans la fenêtre. La fenêtre est ensuite glissée (décalage d'un pas de 1 échantillon). Le signal obtenu après régression est le signal débruité. Les résultats obtenus par cette méthode de débruitage (avec un polynôme de régression d'ordre 2) sont illustrés par la figure 26 et les tables 3 et 4 :

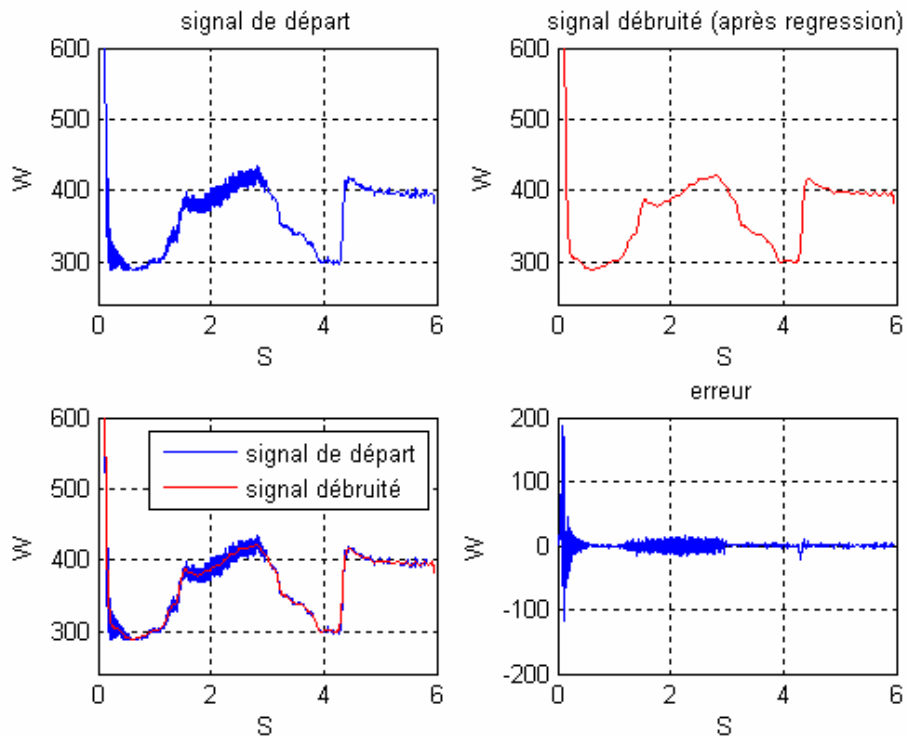


Figure 26 : Résultat du filtrage par régression polynomiale d'ordre 2

Le signal après régression est le signal filtré, la différence entre le signal de départ et le signal de régression (l'erreur) représente le bruit.

On obtient, comme pour le filtre de Kalman, les résultats :

$\bar{\varepsilon}$	σ	P_{ε}
0.5311	16.1865	51.9256

Tableau 3 : Résultats pour un seul signal

Les résultats moyennés sur tous les signaux sont :

$\bar{\varepsilon}$	σ	P_{ε}
0.76	24.66	55.20

Tableau 4 : Résultats du filtrage sur tous les signaux

Cette méthode fournit des meilleurs résultats que l'approche par filtrage de Kalman. On remarque que l'erreur est de moyenne proche de zéro (0.76) et que la variance est plus faible que celle obtenue par la première méthode.

Pour les signaux d'aiguillage, nous avons préféré une méthode classique de filtrage passe-bas filtrage passe-bas de Butterworth, (fonction Matlab butter), fournissant des résultats similaires à l'approche par régression et plus rapide à mettre en œuvre..

2. Segmentation

2.1 Détection du temps de début de manœuvre utile t_0

La segmentation des signaux s'effectue sur une portion de signal dite utile car porteuse de l'information nécessaire pour le diagnostic. L'instant initial de cette portion du signal doit être déterminé.

La localisation de cet instant qui est le début de la phase 1, a été initialement considérée comme fixe. Devant une légère variabilité de la durée de la phase d'appel (phase 0), il a été préféré une détection automatique du t_0 . Le signal utilisé pour cette localisation est le courant moteur, car moins bruité que le signal de puissance. Après débruitage par filtrage passe-bas, le changement abrupt dans le signal est déterminé à l'aide d'un test sur sa dérivée. Pour le test, un seuil sur la dérivée du signal a été fixé afin de ne pas garder la partie du signal où la variation du signal est significative de la phase d'appel. La figure 27 montre un exemple de détection de l'instant t_0 de début de la manœuvre utile.

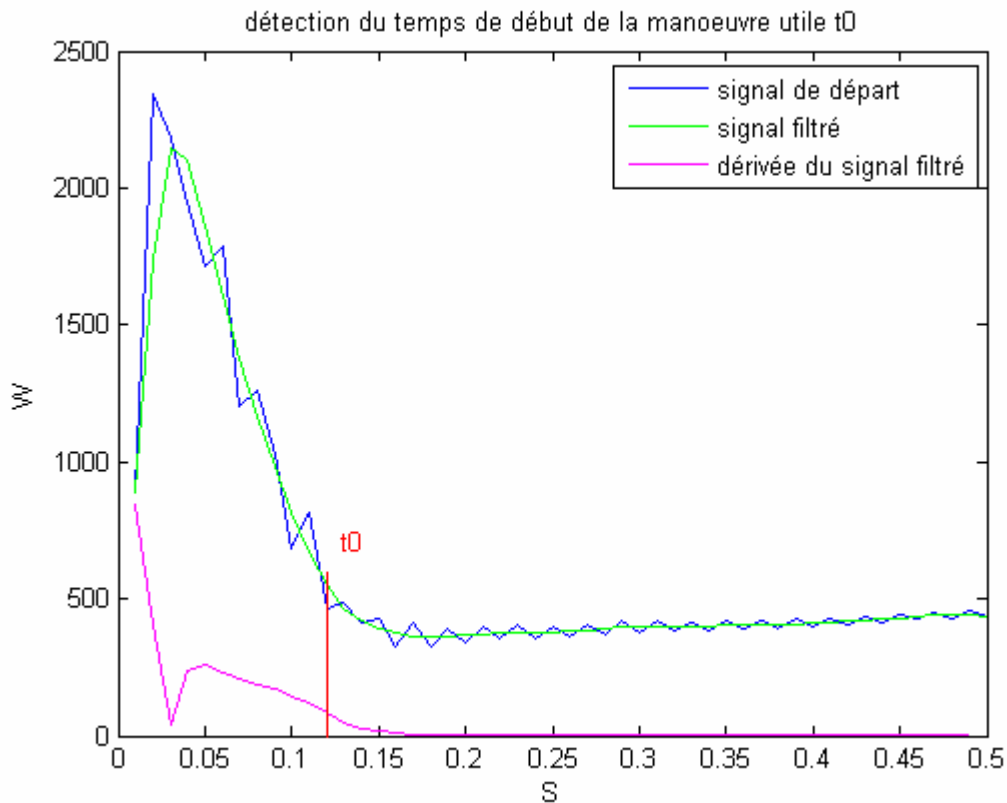


Figure 27 : Détection de t_0

Pour l'ensemble des signaux disponibles, les résultats de détection de t_0 sont les suivants :

- Valeur moyenne de t_0 : 0.1226 S,
- Ecart-type de t_0 : 0.0083 S.

2.2 Définition des bornes des segments

Chaque phase de la manoeuvre est supposée de durée fixe et connue. À partir de l'instant t_0 de début de la phase de Décalage/Déverrouillage, les instants de début de chaque segment sont :

- temps de début de la phase de Décalage/Déverrouillage : t_0 ,
- temps de début de la phase de translation : $t_1 = t_0 + 1.13$ S,
- temps de début de la phase de verrouillage/calage : $t_2 = t_1 + 2.14$ S,
- temps de début de la phase de Friction : $t_3 = t_2 + 1.13$ S.
- temps de fin de manoeuvre : Ce temps est déterminé sur la fin du signal de puissance consommée.

La figure 28 montre une manoeuvre segmentée.

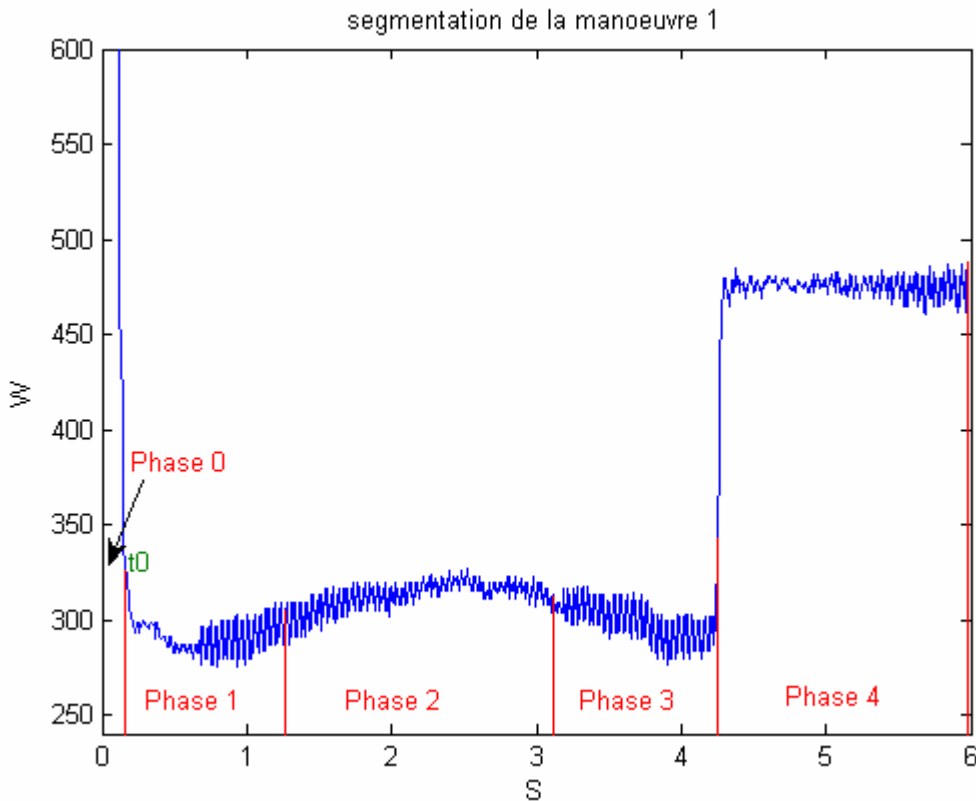


Figure 28 : segmentation d'une manoeuvre

3. Paramétrisation

Comme vu dans le début de ce chapitre, dans un système de reconnaissance de formes, la paramétrisation ou extraction des caractéristiques consiste à calculer un ensemble de paramètres permettant de représenter les mesures de façon aussi précise que possible. Cela permet notamment de travailler dans un espace de dimension plus réduite.

3.1 Choix des paramètres

La paramétrisation du signal se fait par concaténation des paramètres relatifs à chaque phase. Chaque phase est supposée de durée fixe et connue. Le vecteur paramètre pour un signal est finalement le vecteur constitué de plusieurs indicateurs statistiques de chacune des phases. Pour chaque phase, on retient la valeur max, la valeur min, l'espérance, l'écart-type. En plus de ces paramètres, un modèle polynomial du second degré permet de qualifier la « forme » du signal dans la phase (signal symétrique ou non, convexe ou concave...). Ce modèle polynomial obtenu par régression par les moindres carrés s'écrit : $S = \lambda(t - \alpha)^2 + \mu$, où

- α : est le paramètre qui représente la position de l'axe de symétrie du segment,
- λ : L'indice de concavité,
- μ : La valeur au point $t = \alpha$.

3.2 Rappel de la méthode des moindres carrés ordinaires

Soient x_1, x_2, \dots, x_N des mesures issues de la grandeur X , la régression polynomiale non linéaire consiste en représenter les donnée par le modèle suivant :

$$\begin{aligned} x_1 &= a_p t_1^p + a_{p-1} t_1^{p-1} + \dots + a_2 t_1^2 + a_1 t_1 + a_0 + \varepsilon_1 \\ x_2 &= a_p t_2^p + a_{p-1} t_2^{p-1} + \dots + a_2 t_2^2 + a_1 t_2 + a_0 + \varepsilon_2 \\ &\vdots \\ x_n &= a_p t_n^p + a_{p-1} t_n^{p-1} + \dots + a_2 t_n^2 + a_1 t_n + a_0 + \varepsilon_n \end{aligned}$$

Une formalisation matricielle de ce modèle s'écrit :

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} t_1^p & t_1^{p-1} & \cdot & \cdot & t_1 & 1 \\ t_2^p & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ t_{n-1}^p & \cdot & \cdot & \cdot & \cdot & \cdot \\ t_n^p & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} \times \begin{bmatrix} a_p \\ a_{p-1} \\ \cdot \\ \cdot \\ a_1 \\ a_0 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_{n-1} \\ \varepsilon_n \end{bmatrix}$$

Le modèle est donc : $Y = M \times P + \mathcal{E}$ où

- $M = \begin{bmatrix} t_1^p & t_1^{p-1} & \cdot & \cdot & t_1 & 1 \\ t_2^p & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ t_{n-1}^p & \cdot & \cdot & \cdot & \cdot & \cdot \\ t_n^p & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix}$: Matrice du modèle

- $P = \begin{bmatrix} a_p \\ a_{p-1} \\ \cdot \\ \cdot \\ a_1 \\ a_0 \end{bmatrix}$ est le vecteur des paramètres du modèle, et $\mathcal{E} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_{n-1} \\ \varepsilon_n \end{bmatrix}$ le vecteur aléatoire des résidus.

On estime généralement par minimisation de l'erreur quadratique $(X - MP)^T (X - MP) = \sum_{i=1}^N \varepsilon_i^2$. On obtient $\hat{P} = (M^T M)^{-1} M^T \hat{Y}$.

3.3 Paramètres

Le modèle $S = \lambda(t - \alpha)^2 + \mu$ est identifié au modèle polynomial du second degré :

$S = a_2 t^2 + a_1 t + a_0$, pour déterminer les paramètres α , λ et μ . On a :

$$S = \lambda(t - \alpha)^2 + \mu = \lambda t^2 - 2\lambda\alpha + \lambda\alpha^2 + \mu = a_2 t^2 + a_1 t + a_0.$$

Par identification, on obtient les paramètres suivants : $\lambda = a_2$; $\alpha = -\frac{a_1}{2\lambda}$; $\mu = a_0 - \frac{a_1^2}{4\lambda}$.

Le coefficient α est divisé par la taille du segment (nombre d'échantillons) afin d'obtenir une normalisation de ce paramètre par rapport aux différents segments : $\alpha = \frac{\alpha}{\text{taille}(\text{segment})}$. Le

vecteur paramètre pour la phase i est finalement :

$$P_{\text{phase}_i} = [M \ m \ \bar{S} \ \sigma \ \lambda \ \alpha \ a_1 \ a_0],$$

où M est la valeur maximale de la phase considérée, m la valeur minimale, \bar{S} valeur moyenne et σ l'écart-type.

On dispose donc de 32 paramètres pour chaque signal.

3.4 Diagramme de corrélation des variables

La figure 29 montre la représentation des données selon les sept premières caractéristiques

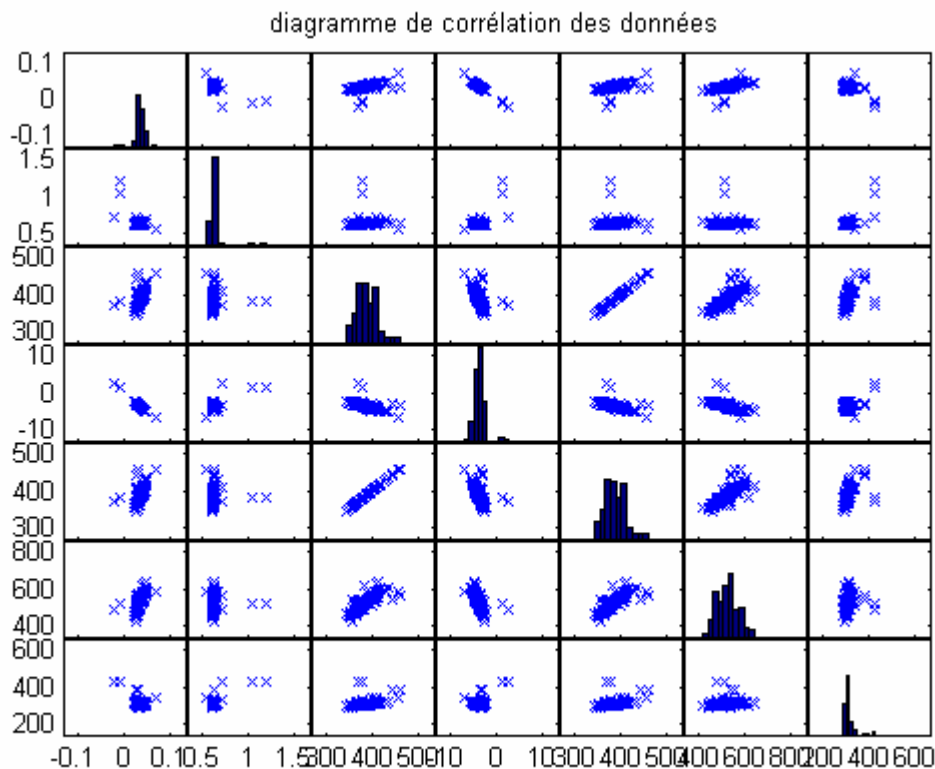


Figure 29 : Représentation des données dans l'espace des caractéristiques

Conclusion

Dans ce chapitre on a présenté les différentes méthodes mises en œuvre pour segmenter et paramétrer les signaux d'aiguillage. La paramétrisation va servir pour la mise en œuvre et l'évaluation du classifieur. La robustesse de celui-ci nous renseignera sur la performance de notre codage. La classification des signaux ainsi que la comparaison de la performance de différents classifieurs seront présentés dans le chapitre suivant.

Chapitre 4. Classification des signaux

Introduction

Cette partie d'étude consiste à choisir une procédure adéquate pour classifier un signal d'aiguillage à partir de son vecteur forme, c'est-à-dire à définir les régions qui caractérisent chaque classe (voir l'exemple de la figure 30).

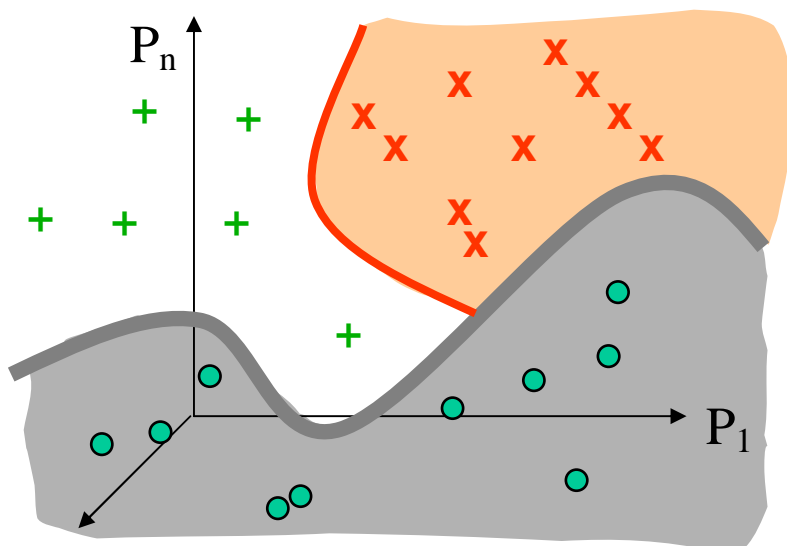


Figure 30 : Exemple de frontières de décision entre les classes.

La décision nous permet de passer de l'espace de représentation, où les données sont représentées selon leurs caractéristiques, à l'espace de décision contenant les différentes classes des données. Cet espace dans notre application comporte trois classes

- Classe 1 : manœuvre normale,
- Classe 2 : manœuvre avec défaut tolérable,
- Classe 3 : manœuvre avec défaut grave.

Dans ce chapitre, plusieurs méthodes (probabilistes et non probabilistes) seront testées pour la classification des signaux. On désignera dans tout ce chapitre les exemples par x_i ($i = 1, \dots, N$) et les classes par C_k ($k = 1, \dots, K$). Les coordonnées de x_i seront notées x_{ir} ($r = 1, \dots, D$).

1. La classification supervisée

1.1 Définition

Dans la classification supervisée, les classes d'appartenance des observations sont connues d'avance. Le but principal de cette méthode de classification est d'apprendre la relation entre les observations et l'information cible (classe), pour :

- Estimer la classe pour de nouvelles observations,
- Comprendre la relation entre les observations et les classes.

1.2 Normalisation des paramètres

L'invariance des mesures en fonction de la translation et du changement d'échelle est souvent souhaitable. Pour cela, nous avons centré et réduit les variables.

2. Mise en œuvre des différentes méthodes de classification supervisée

Dans cette partie, nous appliquons différentes méthodes de classification supervisée aux données d'aiguillage.

Plan expérimental

Le plan expérimental, mis en œuvre sous Matlab, que nous avons adopté, se résume comme suit. On effectue un tirage aléatoire d'observations suivant un pourcentage donné (par rapport à la base complète des 119 observations disponibles), pour former la base d'apprentissage. Les observations n'ayant pas été tirées constituent la base de test. Les différents pourcentages qui ont été considérés pour la base de test sont de 50, 60, 70, 75, 80, 85 et 90 %. Les algorithmes de classification nécessitant un apprentissage de paramètres sont lancés sur la base d'apprentissage afin d'estimer les paramètres de classification, puis les données de test sont classées suivant les paramètres obtenus dans l'étape d'apprentissage. Pour les approches ne nécessitant pas d'apprentissage (approches directes), les données de test sont directement classées. Finalement, le pourcentage de bien classés est calculé. Il faut noter que pour chaque pourcentage de la base d'apprentissage, 25 tirages aléatoires différents ont été effectués, et les pourcentages de bien classés obtenus sur la base de test ont été moyennés sur ces 25 tirages.

2.1 Classification directe

Le principe des méthodes de classification directe est le suivant : à partir des paramètres des données de départ formant un ensemble de vecteurs x_i ayant pour labels associés C_k , l'objectif est de décider de la classe d'appartenance d'un nouveau vecteur x_j en calculant les distances entre x_i et des vecteurs représentant les différentes classes. Parmi les approches de classification directes on peut citer : le classifieur k-nn (k-nearest-neighbor) pour k-ppv (k-plus proches voisins). La discrimination des exemples suivant les différentes classes utilise une distance euclidienne. La classification par k-ppv est une approche non paramétrique car elle ne fait pas d'hypothèse sur la structure des données.

2.1.1 Classification par k-plus proches voisins

Notons par E la base de référence et par F la base de test.

Algorithme

Pour j allant de 1 jusqu'à $Card(F)$ **faire**

Pour i allant de 1 jusqu'à $Card(E)$ **faire**

- Calculer la distance euclidienne entre l'exemple de test et tous les exemples de la base

$$\text{de référence : } d(x_i, x_j) = \sqrt{\sum_{r=1}^D (x_{ir} - x_{jr})^2} .$$

Fin pour

- Déterminer les k vecteurs PP(x) de la base de référence les plus proches,
- Si k=1 : C_j est la classe de l'exemple le plus proche.
- Si k ≠ 1 : C_j est la classe majoritaire dans {PP(X)}

Fin pour

a. Résultats obtenus sur les signaux d'aiguillage

Le tableau suivant présente le taux de bonne reconnaissance en fonction du nombre k de plus proches voisins et du pourcentage d'observations utilisés comme base de référence.

Les résultats sont cités dans le tableau ci-dessous (ce sont les taux de reconnaissance en %)

Nombre de plus proches voisins \ Pourcentage de la base de référence (%)	Nombre de plus proches voisins					
	1	2	3	5	7	11
50	69,49	66,10	59,32	59,32	59,32	64,40
60	79,16	64,58	72,91	64,58	62,5	64,58
70	82,85	65,71	71,42	74,28	80	85,71
75	80	63,33	70	70	73,33	83,33
80	79,16	58,33	66,66	70,83	75	87,50
85	77,77	55,55	61,11	61,11	66,66	77,77
90	83,33	66,66	75	58,33	83,33	75

15	17	19	20	23	26	30
62,71	64,40	64,40	64,40	66,10	66,10	66,10
64,58	64,58	64,58	64,58	66,66	66,66	64,58
85,71	80	77,14	77,14	74,28	80	74,28
86,66	86,66	76,66	80	80	76,66	73,33
79,16	75	79,16	75	79,16	79,16	75
77,77	83,33	72,22	77,77	77,77	88,88	83,33
75	83,33	75	75	75	75	75

Tableau 5 : Résultats de la classification par k-ppv

La figure 31 montre l'évolution du taux de bonne classification en fonction du nombre de plus proches voisins.

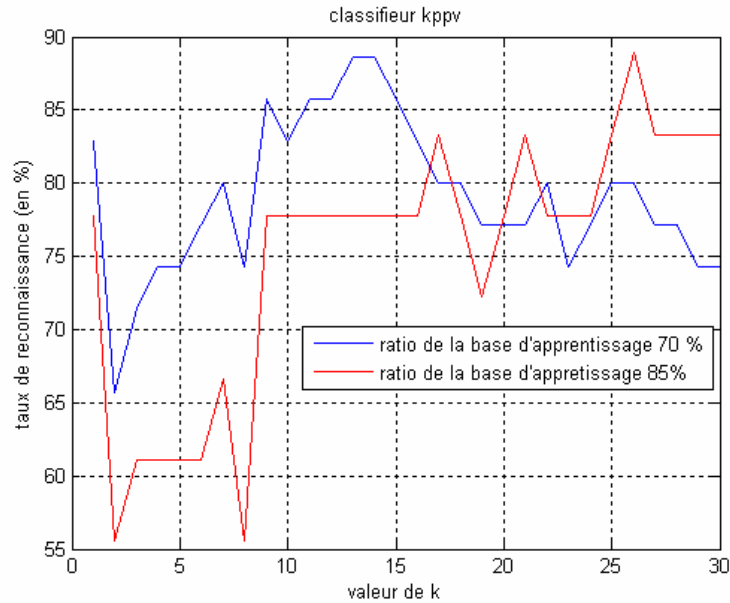


Figure 31 : Variation du taux de reconnaissance avec le nombre de plus proches voisins

La figure 32 montre l'évolution du taux de bonne classification en fonction du pourcentage d'observations dans la base de référence.

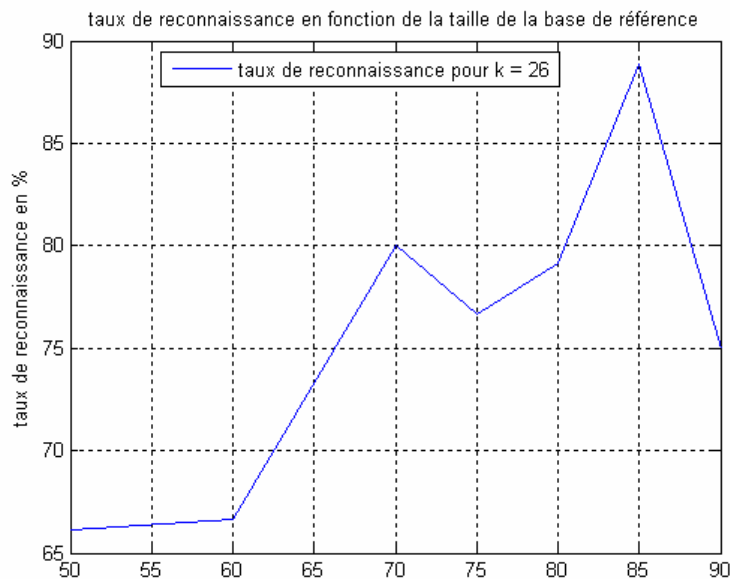


Figure 32 : Variation du taux de bonne classification en fonction de la taille de la base de référence

b. Interprétations

Pour le classifieur kppv, on remarque que le taux de reconnaissance s'accroît avec le nombre de plus proches voisins, puis se stabilise autour d'une valeur (taux théorique de Bayes, [DUBUISSON 01]). Le taux de reconnaissance maximum est de 88.8% pour 26 ppv. Ce comportement observé correspond aux attentes de ce type de classifieur.

La figure 33, illustre la relation entre le nombre des plus proches voisins et le taux de bonne classification. Pour un classifieur 1-ppv, la classe qui sera attribuée à l'exemple de test appartenant à la classe C_1 (en vert), c'est C_2 , puisque l'exemple lui étant le plus proche appartient à cette classe (en rouge). Par contre, si on augmente le nombre de k plus proches voisins, il sera bien classé (dans ce cas de figure, il sera bien classé pour 2-ppv puisque les deux exemples les plus proches appartiennent à la classe C_1 dont appartient cet exemple). D'où l'intérêt d'augmenter la valeur de k , en effet augmenter la valeur de k revient à rendre le classifieur peu sensible au bruit.

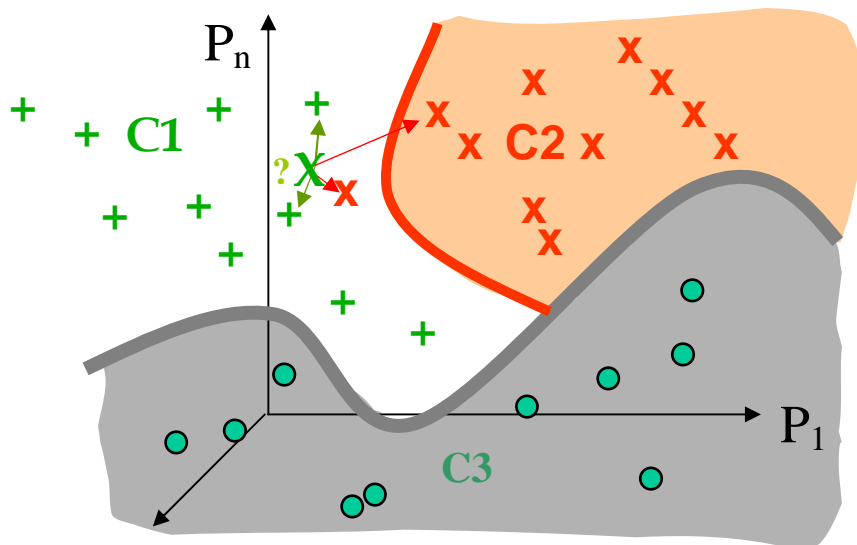


Figure 33 : interprétation de la mauvaise classification k -ppv

Comme pratiquement, du point de vue temps de calcul, augmenter la valeur de k n'augmente quasiment pas le temps de calcul, en effet, il suffit de prendre les k premières classes (déjà calculées pour le cas 1-ppv) et d'en prendre la classe majoritaire.

La figure 32 illustre la variation du taux de reconnaissance lorsque le nombre des plus proches voisins est de 26, en fonction de la taille de la base de référence. Quand la base de référence comporte un faible nombre de points, le taux de bonne reconnaissance est plus faible que lorsqu'elle comprend un plus grand nombre de points. On obtient notamment un taux de bonne reconnaissance de 66.10% à partir d'un ensemble de données de référence formé de 50% des données de la base totale, et un taux de bonne reconnaissance de 88.88% quand l'apprentissage s'effectue sur 85% de la base totale

Le défaut majeur de la méthode des k -ppv reste le problème de son temps de calcul car il nécessite d'effectuer le calcul sur tous les points de la base de référence.

2.1.2 Classifieur Nearest Mean (la moyenne la plus proche)

Le temps de calcul élevé étant le majeur inconvénient d'un classifieur k -ppv, une des variantes du classifieur k -ppv, pour accélérer la classification est le classifieur se basant sur la moyenne la plus proche. En effet, prendre la moyenne de chaque classe de la base de référence au lieu d'en prendre tous les exemples, revient à diminuer considérablement la taille

de la base de référence et par conséquent, la comparaison en terme du plus proche voisin, ne se fera plus avec tous les exemples de référence mais seulement avec les centres des classes; ce qui accélère la classification. Ce classifieur est très voisin de l'analyse factorielle discriminante (AFD).

Cette méthode a été testée dans le cadre de notre application. Elle consiste à affecter un nouvel individu à la classe qui minimise l'écart entre cet individu et les moyennes des classes. L'algorithme de cette méthode se décrit de la manière suivante :

Algorithme

1. Calcul des centres des classes

Pour k allant de 1 jusqu'à K classes de la base de référence **faire**

- $m_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i$: centre de chaque classe de la base de référence.

Fin Pour

2. Classification

Pour i allant de 1 jusqu'à $Card(F)$ **faire**

Pour $k=1,2 \dots K$ Classes de E **faire**

- Calculer la distance euclidienne entre l'exemple de test et tous les centres de classes de la base de référence : $d(x_i, m_k) = \sqrt{\sum_{r=1}^D (x_{ir} - m_{kr})^2}$ où les m_{cr} sont les composantes de m_k .

Fin pour

- Affecter x_i à la classe qui minimise $d(x_i, m_k), 1 < k < K$.

Fin pour

a. Résultats

Les résultats de ce classifieur sont présentés dans le tableau 6 et la figure 34.

% des points de la base de référence	50	60	70	75	80	85	90
Taux de reconnaissance en % (sur la base de test)	61.01	62.50	71.42	66.66	75.00	83.33	75.00

Tableau 6 : Résultats du classifieur de la moyenne la plus proche

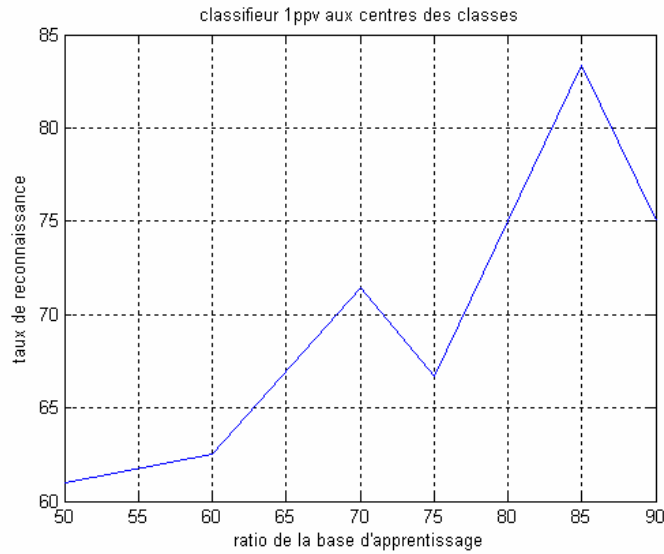


Figure 34 : variation du taux de reconnaissance selon le taux de partage base de référence/base de test

b. Interprétations

On remarque que, plus la distribution des données de référence est homogène, plus la robustesse de ce classifieur s'améliore, ce qui se manifeste bien dans les résultats présentés dans la figure 34 où le taux de reconnaissance augmente avec l'augmentation de la taille de la base de référence. Par exemple, pour un pourcentage de 50 % de la base de référence le taux de bonne reconnaissance est de 61.01 et pour une base de référence de pourcentage 85 % de la base totale, le taux de bonne classification est de 83,33 % (taux maximum atteint).

Cette méthode accélère la classification, Cependant, le fait de considérer le centre de chaque classe comme référence de celle-ci est désavantageux dans le cas où la distribution spatiale des données n'est pas homogène, comme dans le cas où on n'a pas assez d'exemples ou dans le cas où chaque classe de l'ensemble de référence est constituée de sous-classes. Dans ce cas, la variance intra-classes est élevée et la moyenne des points d'une classes peut s'avérer être grossière.

L'exemple de la figure 35 ci-dessus illustre une mauvaise classification d'un exemple de la classe 2 (en rouge), qui est reconnu comme un exemple appartenant à la classe 1 puisque le centre de la classe 1 (m_1) est plus proche que les deux autres centres de classes. Cela est dû à la non homogénéité des données de la base de référence et au fait que la classe 1 est formée de deux sous-classes.

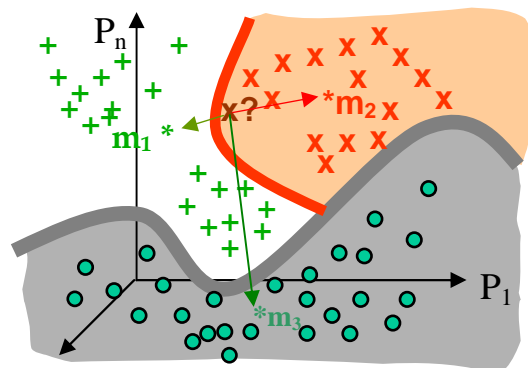


Figure 35: Exemple de mauvaise classification pour la méthode de la moyenne la plus proche

2.2 Approches probabilistes

Le problème des critères métriques est qu'ils ne prennent pas toujours en compte les probabilités a priori et la répartition des données des différents groupes, qui peuvent être très inégales (prévision de défaillance, diagnostic d'un défaut rare ...). De nombreuses erreurs d'affectation sont alors possibles. Il était donc intéressant d'étudier des approches probabilistes.

2.2.1 Rappels sur la classification ou discrimination Bayésienne

Dans la classification Bayésienne les exemples de chaque classe sont supposés être distribués suivant des lois de probabilités. L'affectation d'un nouvel exemple à une classe s'effectue en maximisant, par rapport aux classes indicées par k , les probabilités a posteriori issues de la formule de Bayes :

$$P(C_k / x) = \frac{P(x / C_k) P(C_k)}{P(x)} = \frac{P(x / C_k) P(C_k)}{\sum_{k=1}^K P(x / C_k) P(C_k)},$$

où :

- $P(C_k)$ est la probabilité a priori (en dehors de toutes mesures),
- $P(x / C_k)$ est la densité de probabilité conditionnelle à estimer.

2.2.1.1 Estimation des probabilités a posteriori

L'estimation des probabilités a posteriori nécessite l'estimation des probabilités a priori et des densités de probabilités conditionnelles.

a. Estimation des probabilités à priori

La probabilité a priori $P(C_k)$ est la probabilité de la classe C_k avant d'observer x . En l'absence d'informations particulières, on les suppose égales : $P(C_k) = \frac{1}{K}$ pour les K classes. Si l'échantillon d'apprentissage est représentatif, on utilise l'estimateur fréquentiel $P(C_k) = \frac{N_k}{N}$, où N_k est l'effectif de la classe k et N est l'effectif total.

b. Estimation des densités de probabilités conditionnelles dans le cas gaussien

L'estimation des densités de probabilités conditionnelles de chaque classe peut s'effectuer suivant deux approches :

1. Une approche paramétrique : dans laquelle on fait une hypothèse sur la distribution des données, par exemple une distribution gaussienne, et dans ce cas là, estimer les densités de probabilités revient à estimer l'espérance et la matrice de covariance de chaque classe des données de la base d'apprentissage. Un des estimateurs de ces paramètres est l'estimateur de maximum de vraisemblance qui a de bonnes propriétés

asymptotiques. Cette approche nécessite une bonne connaissance à priori du phénomène.

2. Une approche non paramétrique : dans laquelle on estime la densité des probabilités conditionnelles sans spécifier au préalable la forme sur celle-ci en, on développe donc un algorithme qui converge vers la densité ou la surface idéale quelle qu'elle soit. On distingue parmi ce type d'estimateur, l'histogramme, l'estimateur à base de noyau...

Dans le cas gaussien, que nous avons adopté dans notre application, les densités de probabilités conditionnelles s'écrivent :

$$P(x/C_k) = (2\pi)^{-\frac{D}{2}} (\det \Sigma_k)^{-1} \exp\left(-\frac{1}{2}(x-m_k)^T \Sigma_k^{-1}(x-m_k)\right),$$

où m_k et Σ_k sont respectivement le centre (moyenne) et la matrice de covariance de la classe C_k .

Estimer la densité de probabilité pour chaque classe, c'est estimer m_k et Σ_k pour chaque classe. L'estimation s'effectue par la méthode du maximum de vraisemblance (Maximum Likelihood).

c. Estimation par la méthode du maximum de vraisemblance

Posons $\theta_k = [m_k, \Sigma_k]$. La vraisemblance de θ_k est définie par :

$$L(\theta_k) = \prod_{i=1}^M P(x_i | \theta_k),$$

où M est le nombre de réalisations de x (l'effectif de chaque classe).

Les valeurs estimées $\hat{\theta}_k$ sont prises quand la vraisemblance $L(\theta_k)$ est maximale.

On a $\hat{\theta}_k = \text{ArgMax}[(L(\theta_k))]$

$$\begin{aligned} &= \text{ArgMax}_{\theta_k} (\ln(L(\theta_k))) = \text{ArgMax}_{\theta_k} \left(\sum_{i=1}^M \ln(P(x_i | \theta_k)) \right) \\ &= \text{ArgMax}_{\theta_k} \left\{ \sum_{i=1}^M \ln \left((2\pi)^{-\frac{1}{2}} (\det \Sigma_k)^{-1} e^{-\frac{1}{2}(x_i-m_k)^T \Sigma_k^{-1}(x_i-m_k)} \right) \right\} \\ &= \text{ArgMax}_{\theta_k} \left\{ \sum_{i=1}^M \left(-\frac{1}{2} \ln(2\pi) + \ln((\det \Sigma_k)^{-1}) - \frac{1}{2}(x_i-m_k)^T \Sigma_k^{-1}(x_i-m_k) \right) \right\} \\ &= \text{ArgMax}_{\theta_k} \left\{ \sum_{i=1}^M \left(\ln((\det \Sigma_k)^{-1}) - \frac{1}{2}(x_i-m_k)^T \Sigma_k^{-1}(x_i-m_k) \right) \right\} \\ &= \text{ArgMax}_{\theta_k} \left\{ \ln((\det \Sigma_k)^{-1}) - \sum_{i=1}^M \left(\frac{1}{2}(x_i-m_k)^T \Sigma_k^{-1}(x_i-m_k) \right) \right\} \\ &= \text{ArgMax}_{\theta_k} \left\{ \ln((\det \Sigma_k)^{-1}) - \frac{1}{2}(x-m_k)^T \Sigma_k^{-1}(x-m_k) \right\} \end{aligned}$$

Cette dernière maximisation conduit aux estimateurs [DUBUISSON 01] :

- $\hat{m}_k = \frac{1}{M} \sum_{i=1}^M x_i$,
- $\hat{\Sigma}_k = \frac{1}{M} \sum_{i=1}^M (x_i - m_k)(x_i - m_k)^T$.

La maximisation des probabilités a posteriori par rapport à k revient finalement à celle de

$$\ln(\hat{P}(C_k)) + \ln\left(\left(\det \hat{\Sigma}_k\right)^{-1}\right) - \frac{1}{2}(x - \hat{m}_k)^T \hat{\Sigma}_k^{-1} (x - \hat{m}_k).$$

2.2.1.2 Algorithme de mise en œuvre du classifieur

➤ Phase d'Apprentissage

Pour k=1,2,...,K classes de la base d'apprentissage **faire**

- $\hat{m}_k = \frac{1}{M} \sum_{i=1}^M x_i$,
- $\hat{\Sigma}_k = \frac{1}{M} \sum_{i=1}^M (x_i - m_k)(x_i - m_k)^T$.

Fin pour

➤ Phase de Test

Pour i=1, 2,..., N exemples de la base de test **faire**

Pour k=1, 2 ..K classes de la base d'apprentissage **faire**

- $d(m_k, x_i) = \ln(p(C_k)) + \ln\left(\left(\det \Sigma_k\right)^{-1}\right) - \frac{1}{2}(x_i - \bar{m}_k)^T \Sigma_k^{-1} (x_i - \bar{m}_k)$.

Fin pour

- Affecter x_i à la classe qui maximise $d(m_k, x_i)$.

Fin pour

2.2.1.3 Résultats

Les résultats de la classification Bayésienne dans le cas gaussien sont les suivants :

Ratio base d'apprentissage/base totale (%)	50	60	70	75	80	85	90
Taux de reconnaissance en % (sur la base de test)	52.54	68.75	65.71	66.66	75	72.22	75.00

Tableau 7 : Résultats de la classification Bayésienne

Evaluation du mauvais classement

La figure 36, illustre un cas de classification à deux classes, représentées en 2D. Dans ce cas, deux exemples sont mal classés, le premier qui est l'exemple (en bleu) appartenant à la classe 2 et ayant une probabilité conditionnelle très proche de zéro conditionnellement à la deuxième classe, par contre conditionnellement à la première (dont il n'appartient pas) il a une probabilité conditionnelle de presque 0.5 (voir la figure : exemple au milieu entre 0 et 1), donc par le principe de maximum a postériori, la probabilité a postériori pour cet exemple est maximisée pour la classe 1 dont il n'appartient pas. De même pour l'exemple en vert qui appartient à la classe 1, néanmoins selon la règle du maximum a postériori, il est classé parmi la classe 2. L'erreur de classification augmente quand les classes sont chevauchées.

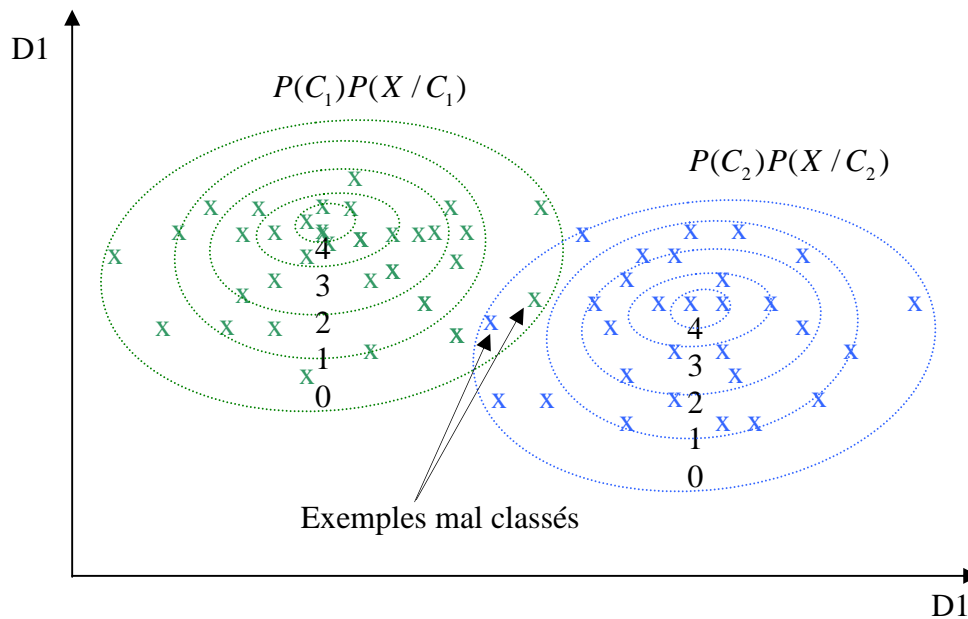


Figure 36 : Cas d'une mauvaise classification dans le cas Bayésien

2.2.1.4 Interprétations

Les résultats du classifieur statistique Bayésien dépendent du nombre d'exemples de la base d'apprentissage. En effet, pour un faible nombre d'exemples, les paramètres statistiques des densités de probabilités à estimer (vecteur moyenne de dimension 32 et matrice de covariance de dimension 23×32 : les moyennes sont correctement estimées, pas les matrices de covariance) conduit à de mauvaises estimations et le taux de reconnaissance se trouve faible. Notamment, pour un ratio de 50 % de la base d'apprentissage par rapport à la base totale, le taux est de 52.54 %).

On observe que plus le nombre d'exemples de la base d'apprentissage augmente plus le taux des bien classés augmente. Le meilleur taux de bonne reconnaissance est de 75 % pour un ratio base d'apprentissage/base totale de 80 %.

Mais globalement, le taux de reconnaissance reste moins élevé que les celui approches de type k-ppv. Cela prouve qu'un modèle avec une seule gaussienne n'est peut être pas suffisant. Une modélisation par mélanges de gaussiennes sera étudiée dans le paragraphe suivant.

2.2.2 Discrimination par mélange de densités gaussiennes

Une extension du classifieur Bayésien gaussien est le classifieur à base de mélange de gaussiennes où chaque densité conditionnelle $P(x | C_k)$ est un mélange de gaussiennes :

$$P(x | C_k) = \sum_{r=1}^{R_k} \pi_{rk} \phi(x | m_{rk}, \Sigma_{rk}),$$

avec,

$$\phi(x | m_{rk}, \Sigma_{rk}) = (2\pi)^{-\frac{D}{2}} (\det \Sigma_{rk})^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x - m_{rk})^T \Sigma_{rk}^{-1} (x - m_{rk})\right),$$

où R_k est le nombre de gaussiennes de la classe C_k , m_{rk} et Σ_{rk} sont respectivement le centre et la matrice de covariance de la $r^{\text{ième}}$ composante gaussienne de la classe C_k . Les proportions π_{rk} des mélanges vérifient :

$$\sum_{r=1}^{R_k} \pi_{rk} = 1, \quad \forall k = 1, \dots, K.$$

L'utilisation des mélanges de gaussiennes en discrimination permet d'obtenir des frontières de décision non linéaires complexes qu'une simple gaussienne ne serait pas en mesure de fournir. La phase de classification s'effectue comme dans le cas d'une gaussienne par maximisation des probabilités a posteriori.

2.2.2.1 Algorithme d'estimation des paramètres des mélanges

L'estimation des paramètres d'un mélange gaussien s'effectue habituellement par la méthode du maximum de vraisemblance. La vraisemblance n'étant pas facile à maximiser dans le cas d'un mélange, l'algorithme EM (Expectation Maximization) [DEMPSTER 77] est généralement utilisé.

Dans le cadre de notre application, compte tenu du nombre très élevé des paramètres à estimer, nous avons imposé la contrainte de matrice de covariance diagonale pour chaque densité gaussienne. Ce modèle particulier permet en outre d'obtenir des estimations plus stables. Le nombre optimal de gaussiennes a été calculé par le critère BIC (Bayesian Information Criterion) [SCHWARZ 78]. Pour la mise en œuvre de l'algorithme, nous avons utilisé le logiciel MIXMOD, implémenté sous Matlab, [MIXMOD] d'estimation de mélanges de gaussiennes.

2.2.2.2 Résultats

Le tableau 8 montre les résultats en termes de bonne classification obtenus avec cette méthode.

% des points de la base d'apprentissage	50	60	70	75	80	85	90
Taux de reconnaissance en % (sur la base de test)	52.54	72.92	77.14	73.33	95.33	94.44	91.67

Tableau 8 : Résultats de la classification par mélanges de gaussiennes

Le tableau 9 donne un exemple du nombre de gaussiennes obtenus pour chaque classe, lorsque 80 % des données ont été utilisées pour l'apprentissage.

Classe	1	2	3
Nombre de gaussiennes	4	2	2

Tableau 9 : Nombre de gaussiennes par classe dans la classification par mélange de gaussiennes

2.2.2.3 Interprétations

On remarque que le taux de reconnaissance augmente avec l'augmentation du nombre d'exemples d'apprentissage. Notamment, on a un taux de bien classés de 52.54 % pour un ratio de 50 % (base d'apprentissage/ base totale). Ce taux s'élève à 95.33 % pour 80 % d'exemples utilisés pour l'apprentissage. En effet l'estimation des paramètres statistiques du mélange dans chaque classe s'améliore avec l'augmentation du nombre d'exemples de chaque classe.

Pour cette méthode de classification, on observe que dans l'ensemble, les résultats sont meilleurs que dans le cas avec une seule gaussienne. Le taux de reconnaissance maximal dans le cas du mélange de gaussiennes est de 95.33 % pour 80 % d'exemples utilisées en apprentissage. Celui-ci est supérieur à celui obtenu par une modélisation par une seule gaussienne (75 % pour un même nombre d'exemples d'apprentissage). Ce résultat correspond à nos attentes sur le fait qu'utiliser un mélange de gaussiennes au lieu d'une seule améliore les frontières de décision.

2.3 Classification par réseau de neurones

2.3.1 Définition d'un réseau de neurones

Un réseau de neurones est un système d'intelligence artificielle consistant en un ensemble de méthodes d'analyse et de traitement des données ayant deux caractéristiques majeures:

- combinaison (réseau) d'éléments simples : les neurones,
- non linéaires (par opposition aux méthodes classiques de l'analyse des données)

Ces méthodes sont vaguement issues de considérations biologiques.

a. Le Neurone artificiel

C'est une unité qui effectue un calcul simple sur ses entrées (x_i), généralement une somme pondérée, dont le résultat est soumis à une transformation souvent non-linéaire $F()$, appelée fonction d'activation.

Le résultat ou valeur de sortie (notée y) est appelé niveau d'activation :

$$y = F\left(\sum_{i=1}^n w_i x_i + w_0\right) = F\left(\sum_{i=0}^n w_i x_i\right) = F(W^T X),$$

où :

- $X = [1, x_1, x_2, \dots, x_n]^T$ est le vecteur d'entrée,

- $W = [w_0, w_1, w_2, \dots, w_n]$ est le vecteur des poids.

Les poids sont des paramètres à déterminer via un algorithme d'apprentissage.

b. Fonction d'activation

Le tableau 10 montre quelques exemples de fonctions d'activations :

Fonction d'activation du neurone k	Sortie
Fonction seuil :	$y_k = \begin{cases} 0 & \text{si } W^T X < 0 \\ 1 & \text{sinon} \end{cases}$
Fonction sigmoïde	$y_k = \frac{1}{1 + \exp(-(W_k^T X))}$
Tangente hyperbolique	$y_k = \frac{1 - e^{-W_k^T X}}{1 + e^{-W_k^T X}}$
Gaussienne	$\phi_k(x) = \exp\left(-\frac{\sum_{i=1}^n (x_i - w_i)^2}{2w_{n+1}^2}\right)$

Tableau 10 : Fonctions d'activation

Un exemple de fonction d'activation sigmoïde est présenté dans la figure 37.

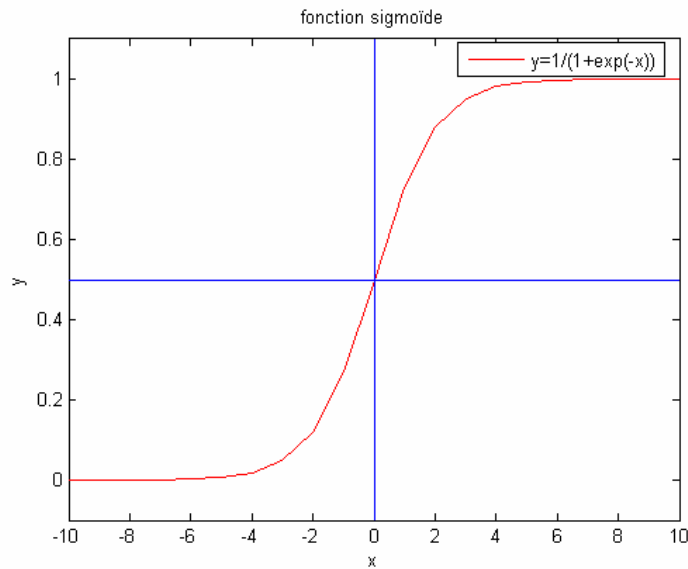


Figure 37 : Fonction d'activation sigmoïde

2.3.2 Perceptron Multicouches (MLP)

a. Le Perceptron :

Un perceptron peut être décrit par le schéma suivant :

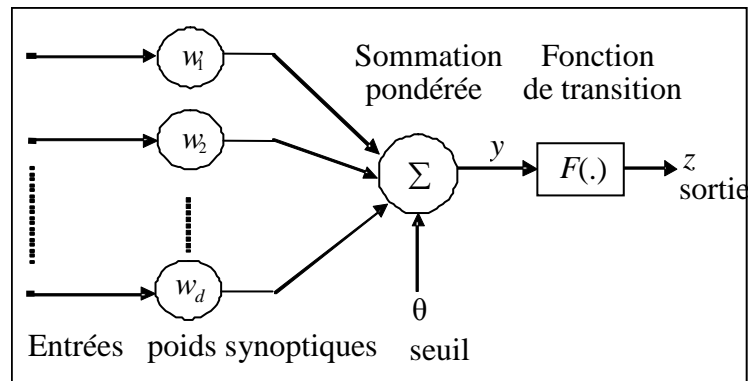


Figure 38 : Perceptron

L'architecture du réseau fixe le schéma d'interconnexion des unités. Chaque connexion est généralement caractérisée par un sens et un poids. On distingue 3 types d'unités :

- les unités d'entrée, activées par un signal d'entrée,
- les neurones de sorties, donnant la réponse du réseau au signal d'entrée,
- les neurones cachés n'ayant pas de connexion avec l'extérieur.

Le dimensionnement du réseau consiste à fixer le nombre de couches cachées et le nombre de neurones par couche. On peut choisir un seul neurone de sortie par classe. La valeur de sortie est la valeur de la classe d'appartenance.

Entrées 1ère couche couche c, c+1, Sorties

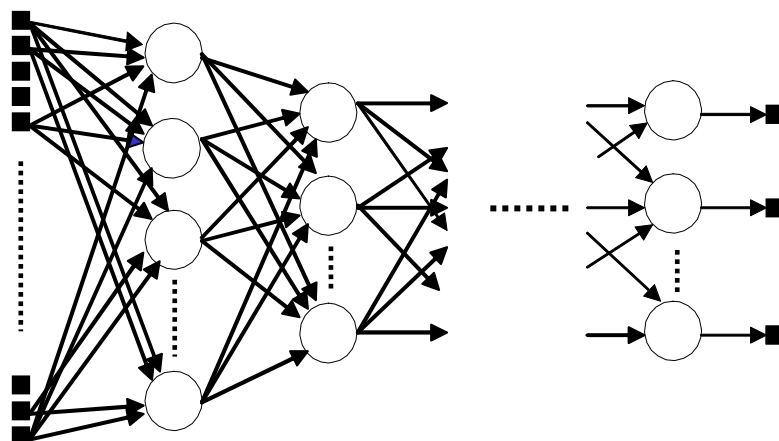


Figure 39 : architecture d'un réseau de neurones type MLP

b. Modèle de réseau utilisé

Le perceptron à deux couches permet de classer des formes non linéairement séparables. Le réseau adopté dans notre application est un perceptron multicouches à une seule couche cachée. La fonction utilisée dans le réseau est une fonction sigmoïde. L'algorithme d'apprentissage est l'algorithme de rétropropagation de gradient classique.

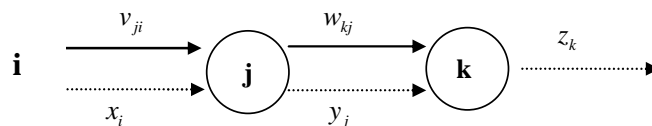
2.3.3 Apprentissage

Le principe d'apprentissage pour un réseau MLP est le suivant [GAS], [Toolbox RdF] :

- Pour chaque vecteur X d'exemples des données d'apprentissage, on propage ces données dans le réseau : propagation du signal de couche en couche :
- On ajuste ensuite les poids associés pour améliorer le processus de classification
- Algorithmes [Toolbox RdF], [Matlab] :
 - Propagation
 - Retro-propagation par Descente de gradient.

Le Perceptron à 2 couches est composé :

- des entrées : m unités (exemples) notées i
- d'une couche cachée : n neurones notés j
- d'une couche de sortie : p neurones notés k



- **Evaluation de l'erreur de classification**

$E_k = \frac{1}{2}(d_k - z_k)^2$ Où d_k est la valeur de la sortie désirée pour ce neurone de la couche de sortie, et z_k la sortie calculée par le réseau

a. Etape de Propagation

Début

Pour chaque neurone j de la couche cachée de 1 à m **faire**

$$s_j \leftarrow \sum_{i=1}^m x_i v_{ji} : \text{entrée du neurone } j \text{ de la couche cachée}$$

$$y_j \leftarrow \frac{1}{1 + e^{-s_j}} : \text{sortie du neurone } j \text{ de la couche cachée}$$

Pour chaque neurone k de la couche de sortie de 1 à n **faire**

$$t_k \leftarrow \sum_{j=1}^n y_j w_{kj} : \text{entrée du neurone } k \text{ de la couche de sortie}$$

$$z_k \leftarrow \frac{1}{1 + e^{-t_k}} : \text{sortie du neurone } k \text{ de la couche de sortie}$$

Fin Pour

Fin Pour

FIN

b. Rétropropagation par descente de gradient

• Algorithme itératif basé sur des méthodes de descente de gradient [Toolbox RdF], [Matlab].

• Principe :

- Initialisation aléatoire des poids
- Recherche du minimum de la fonction E (erreur) dans l'espace des poids pour tous les exemples d'apprentissages par annulation des dérivées partielles :

$$\frac{\partial E}{\partial w_{ij}} = 0, \forall i, j$$

- Modification des poids :

$$w'_{ij} = w_{ij} + \Delta w_{ij} \text{ avec } \Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}},$$

où η est le pas de descente choisi suffisamment petit. On a donc :

$$w'_{kj} = w_{kj} + \Delta w_{kj} \text{ avec } \Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}} = -\eta \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial t_k} \frac{\partial t_k}{\partial w_{kj}},$$

$$v'_{ji} = v_{ji} + \Delta v_{ji} \text{ avec } \Delta v_{ji} = -\eta \frac{\partial E}{\partial v_{ji}} = -\eta \sum_{k=1}^p \frac{\partial E}{\partial z_k} \frac{\partial z_k}{\partial t_k} \frac{\partial t_k}{\partial y_j} \frac{\partial y_j}{\partial s_j} \frac{\partial s_j}{\partial v_{ji}}.$$

c. Fonction Descente-gradient

Début

Pour chaque neurone k de la couche de sortie **faire**

Pour chaque neurone j de la couche cachée **faire**

$$\Delta w_{kj} = -\eta (d_k - z_k) (1 - z_k) z_k y_j$$

$$w_{kj} \leftarrow w_{kj} + \Delta w_{kj}$$

Pour chaque neurone j de la couche cachée **faire**

Pour chaque unité i de la couche d'entrée **faire**

$$\Delta v_{ji} = -\eta x_i (1 - y_j) y_j \left(\sum_{k=1}^p (d_k - z_k) (1 - z_k) z_k w_{kj} \right)$$

$$v_{ji} \leftarrow v_{ji} + \Delta v_{ji}$$

Fin

d. Illustration de la Descente de gradient

L'erreur globale commise sur tous le réseau est $E(\Omega) = \| D^T - X^T \Omega^T \|^2$, où

- Ω est la matrice des vecteurs de poids w_{ij} de tout le réseau,
- D^T est le vecteur des sorties désirées d ,
- X^T est l'ensemble des entrées.

La minimisation de cette erreur repose sur une méthode de descente du gradient. On calcule le gradient de l'erreur par rapport au poids et on effectue une correction dans le sens inverse [GAS].

La figure 40 illustre le principe de descente de gradient.

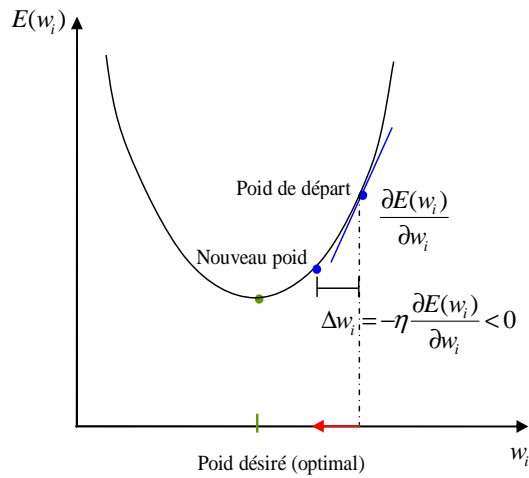


Figure 40 : Principe de la descente du gradient

2.3.4 Résultats

Les résultats de ce classifieur sont présentés dans le tableau 11, 12 et les figures 41 et 42.

% de la base d'apprentissage \ Nbre de neurones cachés	Nbre de neurones cachés									
	3	5	6	7	8	9	10	11	12	
50	66.10	64.40	62.71	67.79	66.10	66.10	66.10	66.10	67.79	
60	66.66	66.66	66.66	68.75	79.16	66.66	68.75	68.75	66.66	
70	68.57	77.14	85.71	85.71	82.85	80	88.5714	82.85	82.85	
75	76.66	76.66	80	83.33	86.66	80	83.3333	83.33	83.33	
80	70.83	83.33	83.33	70.83	83.33	70.83	87.5	79.16	83.33	
85	77.77	77.77	77.77	77.77	83.33	72.22	77.7778	83.33	77.77	
90	75	75	75	75	83.33	75	75	75	75	

	13	14	15	20	25	30
64.40		66.10	66.10	66.10	64.40	72.88
64.58		70.83	66.66	66.66	64.58	72.91
80		88.5714	74.28	88.57	82.85	77.14
90		76.66	73.33	83.33	73.33	76.66
87.5		83.33	75	83.33	79.16	79.16
88.88		77.77	77.77	83.33	77.77	77.77
83.33		75	75	75	75	75

Tableau 11 : Résultats du classifieur neuronal

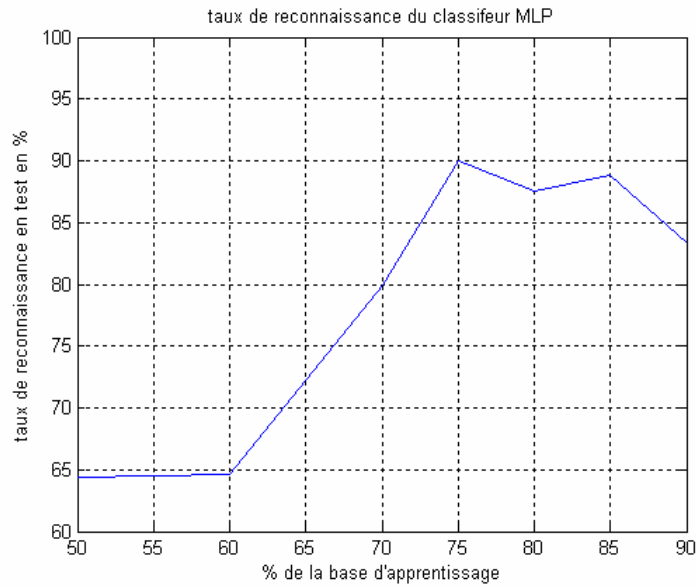


Figure 41 : Variation du taux de bonne classification en fonction de la taille de la base d'apprentissage.

Pour un taux base d'apprentissage/base totale de 75%, les résultats sont les suivants :

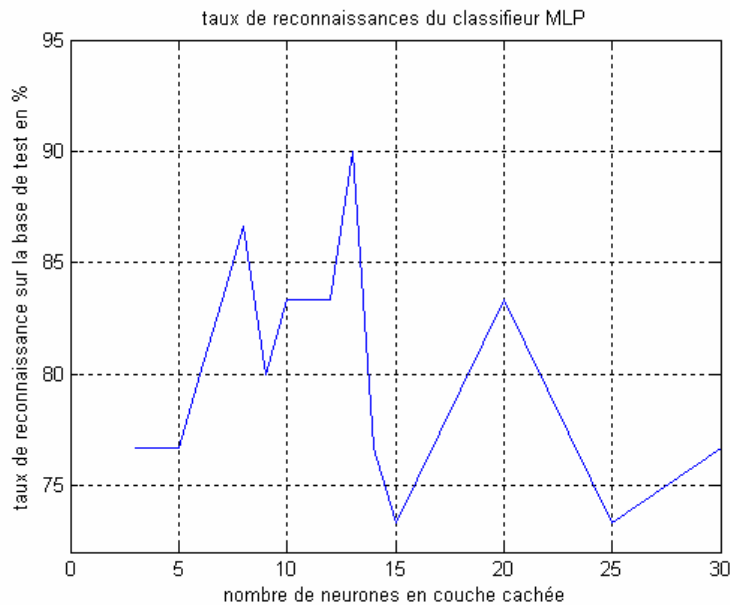


Figure 42: Variation du taux de reconnaissance en fonction du nombre de neurones dans la couche cachée

- Matrice de confusion (75%, 13 cellules en couche cachée)

	C1	C2	C3
C1	6	1	0
C2	1	5	0
C3	0	1	16

Tableau 12 : Matrice des confusions

2.3.5 Interprétations

Pour ce classifieur on remarque que le taux de bonne reconnaissance augmente avec le nombre d'exemples dans la base d'apprentissage. Quand le nombre de cellules en couche cachée est fixe (dans le cas de 13 par exemple), pour un taux de partage (base d'apprentissage/base totale) de 50 % le taux de bien classés est de 64.40 %. Ce taux s'élève à 90 % (taux maximal) quand le pourcentage des exemples d'apprentissage est de 75 % du nombre d'exemples total. Cependant, quand le nombre d'exemples d'apprentissage devient trop élevé (à partir de 80% du nombre total d'exemples), le taux de bonne reconnaissance décroît, cela s'explique par le sur-apprentissage du réseau.

Le taux de bien classés varie également en fonction du nombre de cellules en couche cachée comme le montre la figure 42. En effet, pour un nombre d'exemples de la base d'apprentissage fixé (par exemple 75 % du nombre total des exemples), le taux de bien classés obtenu avec 3 neurones en couche cachée est de 64.40 %. Ce taux s'élève à 90 % quand on a 13 neurones en couche cachée. Ce taux a tendance à décroître quand le nombre des cellules cachées dépasse 20. Cette décroissance peut être attribuée au sur-apprentissage.

La figure 43 illustre le problème de dimensionnement d'un réseau de neurone.

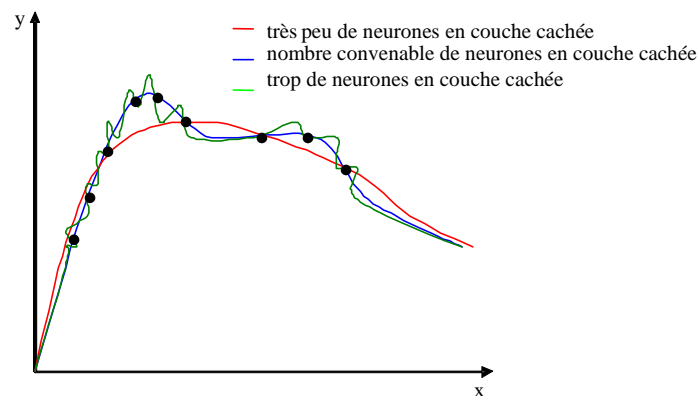


Figure 43 : problèmes de dimensionnement d'un réseau de neurones

La fonction (qui lie les entrées x aux sorties y) à apprendre par le réseau est représentée sur cette figure (figure 43) par les points en noir. Quand le nombre de cellules en couche cachée est insuffisant, le réseau n'arrive pas à bien modéliser les classes et par conséquent l'erreur de classification est grande. Dans le cas où le nombre de cellules en couche cachée est trop élevé, le réseau « apprend trop » les poids et le taux de reconnaissance en généralisation baisse, c'est le phénomène du « sur-apprentissage ».

Conclusion

Dans ce chapitre on a mis en œuvre différentes méthodes de classification pour le diagnostic. Les résultats obtenus dans les cas des approches directes (k-ppv et méthode de la moyenne la plus proche), et dans le cas de l'approche probabiliste avec modélisation par une seule gaussienne sont moins satisfaisantes que ceux obtenus par l'approche probabiliste par mélange de gaussiennes et par le classifieur neuronal.

La figure 43 illustre les résultats obtenus par les différentes approches, en montrant la variation du taux de reconnaissance en fonction du nombre d'exemples d'apprentissage.

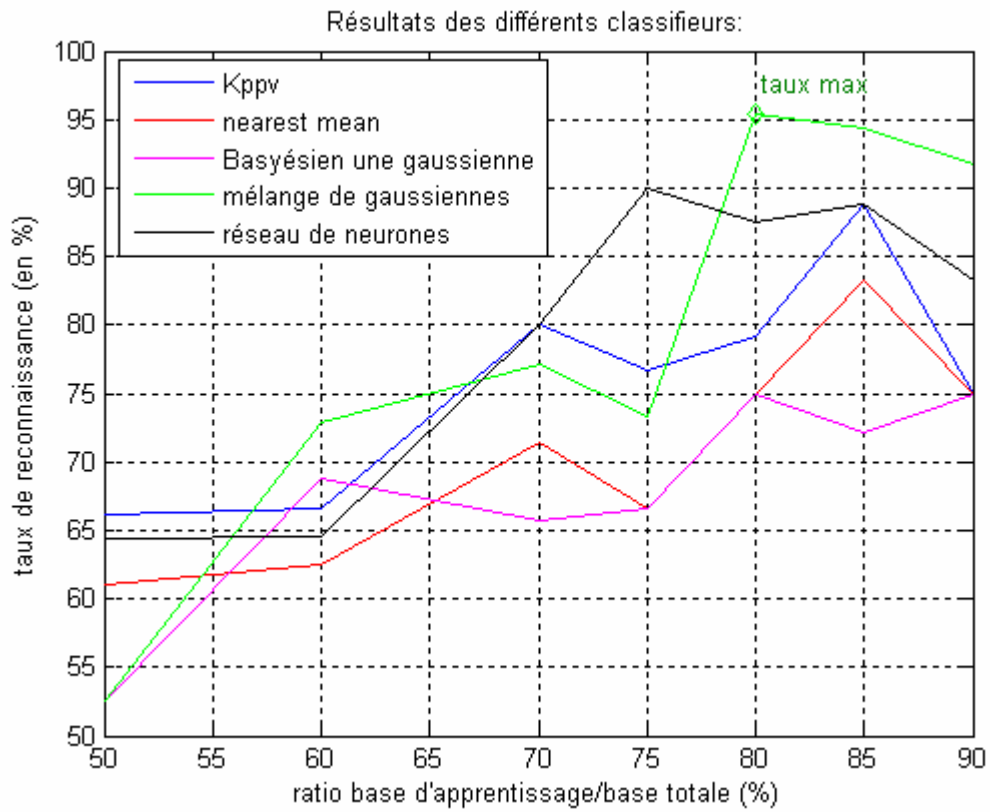


Figure 43 : Résultats des différentes approches de classification

Conclusion Générale

Le travail présenté dans ce mémoire a consisté à développer une méthodologie de diagnostic d'un composant de l'infrastructure ferroviaire, le mécanisme d'aiguillage des rails. Cette tâche a été réalisée à partir de signaux mesurés sur les moteurs d'aiguillage, couvrant un large panel de défauts. Une approche basée sur la reconnaissance des formes a ainsi été proposée. L'avantage de cette approche est qu'elle ne nécessite aucun modèle physique du système à diagnostiquer et qu'elle peut aisément s'adapter à des technologies d'aiguillage différentes.

Après avoir défini un espace de représentation des signaux ad-hoc, nous avons mis en œuvre différentes approches de classification : des approches directes (la méthode des k-plus proches voisins, et une méthode voisine de l'analyse factorielle discriminante), des approches probabilistes et une approche neuronale. Les deux approches probabilistes utilisées s'appuient sur la règle de décision Bayésienne, l'une se basant sur un modèle gaussien pour chaque classe, et l'autre sur un modèle de mélange gaussien pour chaque classe. La méthode neuronale utilise un perceptron multicouche à une couche cachée.

Les résultats obtenus avec ces approches ont mis en évidence de très bonnes performances des approches probabilistes à base de mélange de lois gaussiennes et neuronale, par rapport aux autres approches (classification directe et approche probabiliste avec une seule gaussienne).

A court terme, les perspectives de ces travaux se focaliseront sur l'optimisation de l'espace de représentation, relativement à sa pertinence pour la classification. Ils tenteront également d'intégrer des données de type déphasage courant-tension pour aider au diagnostic. La définition d'une signature de référence pour chaque aiguillage permettra également une approche différentielle possiblement intéressante, surtout d'un point de vue industriel.

A plus long terme, les perspectives de ce travail sont centrées sur deux aspects principaux :

- le suivi de point de fonctionnement lors de manœuvres successives organisées en séquence, avec des approches markoviennes intégrant les connaissances sur les vitesses de dégradation
- le problème du multi labels, où un point de fonctionnement peut correspondre à 2 défauts simultanément.

Bibliographie

- [DEBIOLLES 07] Alexandra DEBIOLLES, « Diagnostic et suivi de point de fonctionnement. Application à l'analyse des circuits de voie ferroviaires », Thèse de doctorat, Université de Technologie de Compiègne, Février 2007.
- [DEMPSTER 77] A. P. Dempster, N. M. Laird, and D. B. Rubin, « Maximum likelihood from incomplete data via the EM algorithm (with discussion) ». Journal of the Royal Statistical Society B, pages 39, PP. 138, 1977.
- [DOC] Guide d'exploitation des relevés graphiques de consommation des mécanismes de manœuvre des aiguilles. Rapport rédigé par Mr. Marc Antoni.
- [DUBUISSON 01] Bernard DUBUISSON, « Diagnostic, intelligence artificielle et reconnaissance des formes », Paris, Hermès Science Publications, ISBN 2-7462-0249-2 p.117-120 135, 2001.
- [GAS] Bruno GAS, « Modèles connexionnistes, Apprentissage et Fusion d'information » <http://www.edu.upmc.fr/sdi/signal-images/BGas/>.
- [INRETS] Site web INRETS : <http://www.inrets.fr>.
- [MIXMOD] <http://www-math.univ-fcomte.fr/mixmod>.
- [MATLAB] Signal Processing Toolbox, Matlab 7.3.0.
- [OUKHELLOU 97] Latifa OUKHELLOU, « Paramétrisation et Classification de Signaux en Contrôle Non Destructif. Application à la Reconnaissance des Défauts de Rails par Courants de Foucault », Thèse de doctorat, Université Université PARIS-SUD, Centre d'ORSAY, Avril 1997.
- [PEDREGAL 04] Pedregal D.J.; Garcia F.P, Schmid F « RCM² predictive maintenance of railway systems based on unobserved components models », Elsevier, Reliability Engineering and System Safety, Volume 83, Number 1, pp. 103-110(8), January 2004.
- [REGIS 01] Lengellé, Régis (sous la direction de), « Décision et reconnaissance de formes en signal », Paris, Hermès Science Publications, ISBN 2-7462-0446-0, p. 200-203, 2004.
- [ROBARDET] CE LINE ROBARDET, « classification supervisée » : <http://prisma.insa-lyon.fr/scrobardet>
- [SCHWARZ 78] Schwarz, « Estimating the dimension of a model ». Annals of Statistics 6(2):461-464, 1978
- [Toolbox RdF] Toolbox RdF 3.1.10 <http://www.edu.upmc.fr/sdi/signal-images/BGas/Master-SSIR/RSI+TSSI/ModelsConnexionnistes>.