

Une méthodologie d'apprentissage automatique pour l'aide à la décision en contexte d'industrialisation

A machine learning methodology for decision support in a Computer Aided Manufacturing context

Emeric Ostermeyer¹, Christophe Danjou², Alexandre Durupt¹, Julien Le Duigou¹

¹ Laboratoire Roberval FRE UTC/CNRS 2012, Université de technologie de Compiègne, France, {emeric.ostermeyer; alexandre.durupt; julien.le-duigou}@utc.fr

² Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation, Polytechnique Montréal, Montréal, Canada, christophe.danjou@polymtl.ca

RÉSUMÉ. L'évolution des techniques de fouilles de données, ainsi que l'accroissement des capacités de stockage et de calcul suscite, dans tous les domaines, un intérêt pour les données produites. En ce sens, le domaine manufacturier ne fait pas exception. Au vu de la quantité de données créées lors de l'écriture des différents programmes à jouer sur les machines à commande numérique, l'application de techniques de fouille de données afin de capitaliser les connaissances d'industrialisation est considérée. Les travaux présentés dans ce papier concernent la structuration d'un système d'Industrialisation à Base de Connaissances, capable d'apporter à un programmeur une aide à la décision, à partir d'un corpus de documents relatifs à l'usinage de pièces déjà réalisées, et ainsi de l'assister dans la réalisation d'une nouvelle pièce. Le système utilise des techniques de fouille de données pour extraire cette information et la délivrer au programmeur.

ABSTRACT. The evolution of data mining techniques, as well as the increase in storage and computing capacity, in all areas, is generating interest of the data produced. In this way, manufacturing is no exception. Given the amount of data created when writing the various programs to be played on CNC machines, the application of data mining techniques to capitalize industrialization knowledge is considered. This paper concerns the structuring of an Industrialization Knowledge Base system, able to provide a programmer decision support, based on a corpus of documents relating to the machining of parts produced in the past, and thus assisting him in the production of a new part. The system uses data mining techniques to extract this information and deliver it to the programmer.

MOTS-CLÉS. Fabrication Assisté par Ordinateur, Aide à la décision, Apprentissage automatique, Système à base de connaissances.

KEYWORDS. Computer Aided Manufacturing, Decision support, Machine learning, Knowledge Based Engineering.

1. Introduction

Dans un contexte manufacturier où sont encouragées la flexibilité de la production, et la diversité des produits, il est nécessaire de réduire les coûts et le temps de développement des étapes de la production, et donc des programmes d'usinage. Les avancées récentes dans les domaines du traitement et de l'exploitation des données, associées à l'importante production de ces dernières, permettent d'imaginer une aide à l'élaboration de programmes d'usinage basée sur des techniques de fouille de données – Définies par [FAY 96] comme des méthodes cherchant à extraire des tendances mathématiques (ex : régression), ou déterminer des similitudes à partir de données.

Malgré la constatation de [LU 90], selon laquelle le domaine de la fabrication produit beaucoup de données, mais peu significatives – La donnée de fabrication intéressante peut être noyée dans la masse des données produites – de nombreux acteurs de la fabrication considèrent comme prometteuses les techniques de fouille de données, pour améliorer leurs procédés. [MAN 16] utilisent ainsi des techniques d'apprentissage automatique pour améliorer la performance d'une ligne de production d'une usine Bosch, à partir de données rendues publiques sur une plateforme de

compétition dédiée à l'application de ces techniques. De même, les auteurs considèrent la possibilité d'assister l'écriture de programmes de machines à commande numérique (CN).

La première étape de la découverte de connaissance à partir de données est, d'après [FAY 96], la sélection des données à partir desquelles la connaissance peut être déduite. Dans cette optique, le domaine de la fabrication peut s'avérer problématique, car les données sont souvent distribuées à travers une hétérogénéité de formats, depuis des dessins de définition à des ordres de fabrication, en passant par des modèles géométriques ou des programmes de CN. Tous ces formats sont destinés à être utilisés pour une application précise, à destination d'un corps de métier spécifique. Ils décrivent donc une partie bien précise des informations nécessaires à l'écriture d'un programme, en utilisant un vocabulaire spécifique. Ainsi, une étape de traduction et d'intégration est nécessaire avant de pouvoir exploiter ces données.

Depuis 2003, la norme STEP-NC [ISO 03] – une norme ISO ayant pour objectif de décrire la pièce et le procédé la produisant, dans le contexte de l'usinage à commande numérique – est vu comme une solution potentielle à ce problème d'hétérogénéité des formats et des expertises. L'échange de connaissances liées aux CN est ainsi rendu possible, grâce à un format centré sur l'entité géométrique. Ce format de fichier n'est cependant pas pensé comme une structure facilement requêttable, qui permettrait par exemple de retrouver facilement les paramètres de coupe utilisés pour un outil spécifique.

Le système d'information d'entreprise contient les données relatives aux pièces déjà réalisées [HUE 11], mais ne permettent pas une recherche précise sur une opération donnée. La chaîne numérique ne permet qu'un transfert d'information vers l'aval.

Afin de permettre la capitalisation des connaissances contenues dans le système d'information d'entreprise, ainsi que de permettre leur utilisation par le programmeur, une étude bibliographique sur les systèmes d'aide à la décision pour la production ainsi que les applications réalisées à partir de techniques de fouille de données est synthétisée en partie 2. La partie 3 présente notre proposition, une méthodologie d'intégration des techniques de fouille de données à partir des données récupérées et indexées pour l'aide à la décision lors de la phase d'industrialisation. La partie 4 présente un cas d'étude sur l'usinage de poches utilisant les réseaux de neurones et la partie 5 conclut ce papier.

2. Le système d'aide à la décision

Avec comme objectif d'utiliser les données déjà présentes dans le système d'information d'entreprise, le système d'aide à la décision doit permettre à l'utilisateur de requêter des informations liées à des décisions antérieures. Pour ce faire, le concept de système à base de connaissances a été considéré.

[LAR 11] définit ces systèmes comme une technologie dont le but est de capturer et réutiliser la connaissance, pour l'automatisation de tâches routinières, et le support de projets multidisciplinaires. Cette définition sied à un outil permettant de retrouver des informations de procédés dans une large base. En effet, la conception et la fabrication d'une pièce impliquent de nombreux corps de métier, et l'écriture de programmes d'usinage implique souvent des tâches répétitives ou peu variées. La partie à venir traitera de l'application des systèmes à base de connaissances à l'industrie manufacturière.

2.1. Les systèmes à base de connaissances pour l'industrialisation

L'utilisation de systèmes à base de connaissances pour assister la planification n'est pas neuve, les premières tentatives datant des années 1980 [ALT 89]. Ces applications ont d'abord été pensées

comme un support des systèmes experts. [ARE 00] utilisent un système à base de connaissances – nommé EXCATS – où la base est écrite par des experts, et les règles qu'elle contient sont appliquées au choix d'outils de tournage. Ce système requiert, pour sa conception, l'implication totale des experts du domaine, et ne peut résoudre que des tâches prédéterminées. Ces systèmes ont, depuis, inclus des données réelles pour la validation et la consolidation du modèle défini par les experts, comme dans les travaux de [TOR 13], où des modèles de régression sont créés a priori, et ensuite adaptés aux données réelles. Le modèle donnant les meilleurs résultats est sélectionné, et considéré comme le meilleur possible. Il est ensuite utilisé comme fonction objectif d'un algorithme évolutionnaire.

L'approche à base de cas met l'accent sur les cas passés. Cette approche considère une base de cas déjà réalisés comme une base de connaissance, et le système utilise une métrique prédéfinie pour évaluer la distance d'un cas nouveau à un cas en base. [CAO 13] appliquent cette approche à la sélection de paramètres de coupe. Ils développent une distance pondérée, considérant les paramètres du matériau et du procédé, associés à un poids variable, afin de permettre un réglage fin de cette métrique. La proposition issue d'un outil utilisant une approche à base de cas prend en compte l'expérience passée, et est à même d'évoluer au fil du temps, mais ne propose qu'une quantité finie de solutions, identiques à des cas déjà vus. Une base de cas conséquente permet de proposer des solutions plus fines.

Le système proposé ci-dessus nécessite une base de données capable de décrire finement les paramètres de l'usinage. En supposant la nécessité d'une base dédiée, les auteurs ont considéré le problème de la récupération et de la mise en base des données. En effet, un système à base de connaissances ayant pour vocation de traiter l'ensemble du processus de planification aurait besoin d'une base comprenant l'intégralité des données de l'entreprise relatives à l'activité d'usinage. Ainsi, [FAN 08], afin de traiter les problèmes d'interopérabilité liés à l'utilisation simultanée de plusieurs systèmes à base de connaissances, proposent une unification des différents domaines au sein d'un même standard. [BER 08] détaillent l'intérêt de lier ces systèmes à un service de gestion du cycle de vie produit (PLM), en soulignant les problèmes liés à la gestion de formats propriétaires.

Cette courte revue souligne un manque de systèmes d'aide à la décision traitant les problèmes de planification de manière générale, et capable de s'adapter aux changements du domaine manufacturier. De plus, l'expertise provient souvent de modèles construits a priori, alors que la quantité importante de données contenue dans les systèmes d'information d'entreprise appelle à une utilisation a posteriori.

Une fois les données représentées dans un format qui permet une indexation aisée, il devient possible d'en explorer les utilisations intéressantes. Les techniques de fouille de données semblent intéresser de plus en plus les acteurs de la fabrication. Ainsi, la seconde partie de cette section sera dédiée à une rapide revue des différentes applications de ces techniques dans notre domaine d'intérêt.

2.2. La fouille de données, et ses applications dans le domaine de la fabrication

La fouille de données, et plus encore son application aux procédés de fabrication, est un domaine encore jeune. Son essor est dû à des capacités de stockage de données de plus en plus importantes, ainsi que du gain d'intérêt récent pour les techniques d'apprentissage automatique. [GE 17] a ainsi listé les applications de la fouille de données dans les industries du procédé, et a détaillé le fonctionnement des différentes techniques passées en revue. La majorité des papiers traitent de la surveillance industrielle, de la classification de défauts et de la prédiction de qualité. [WUE 16] soulignent les avantages résultant de l'application de techniques apprenantes dans le domaine manufacturier, tout en détaillant les difficultés qui peuvent en découler. Ainsi, la donnée y est

souvent lacunaire, et des problèmes de sécurité peuvent entraver l'accès à des données complètes. De plus, elles sont souvent déséquilibrées – il y a typiquement plus de données de production « normales » que de données sur une production « anormale », défectueuse. Enfin, la quantité de techniques apprenantes disponibles est souvent un frein à leur application, dans un domaine où elles sont méconnues. Il existe cependant des applications récentes de ces techniques dans la littérature.

[JOS 08] propose une méthode de sélection de posages pour la réalisation d'une pièce, en utilisant des réseaux de neurones artificiels, et le séquençement d'opérations via un algorithme génétique. [DEN 14] ont étudié la possibilité d'appliquer la méthode des k-moyennes à la sélection de procédés d'usinage, afin de prédire l'état de surface obtenu, tout en donnant des renseignements sur le débit copeau, une métrique importante pour évaluer le coût d'une opération d'usinage. Une machine à vecteur support est comparée à une analyse statistique dans [DEN 16a] et [DEN 16b] afin de proposer des paramètres de coupe, tout en prenant en compte la déviation de forme qui résultera de l'opération. La machine à vecteur support donne les résultats les plus satisfaisants. [ROJ 17] a proposé une méthode de sélection de machine, d'outil et de paramètres de coupe afin de réaliser une opération, à l'aide de trois réseaux de neurones. [HAN 16] détaille l'utilisation de machines à vecteur support pour prédire l'usure d'un outil, et appliquer les corrections nécessaires aux programmes. L'accent est mis sur le prétraitement des données, afin d'obtenir une prédiction viable. Enfin, [PER 14] ont testé plusieurs méthodes de fouille de données pour la résolution d'un problème de diagnostic de défaut, en utilisant des données de simulation. Le papier se conclut en indiquant que – dans ce cas – l'analyse statistique propose souvent une des, sinon la, classification la plus appropriée.

Cette revue des applications de techniques de fouille de données à l'usinage souligne la multiplicité des pistes d'amélioration possibles. De même, les discussions avec les partenaires industriels (voir remerciements) font émerger de nombreux besoins – de la définition de couples outil-matière à des conseils sur le séquençage d'opérations dans des cas complexes. La prochaine partie propose une méthodologie d'application de techniques de fouille afin de répondre au mieux aux besoins industriels.

3. Proposition d'un système d'aide à la décision pour l'industrialisation

Cette section introduit l'architecture générale d'un système à base de connaissances capable de faire face aux problèmes susmentionnés, et prenant en compte les avertissements trouvés dans la littérature scientifique.

La première partie décrit la partie capitalisation de connaissance, permettant de retrouver et restituer la connaissance en se basant sur une base ontologique et le standard STEP-NC. La seconde partie décrit une méthodologie à adopter pour choisir la bonne approche pour mettre en place le moteur d'inférence, afin de choisir parmi les nombreuses approches de fouilles de données (statiques et d'apprentissage automatique) possibles.

3.1. Retrouver et restituer la connaissance : *OntoSTEP-NC*

Afin de proposer un outil flexible au programmeur CN, et grâce à la quantité et qualité des données contenues dans le système d'information d'entreprise, un système à base de connaissances basé sur les données est considéré. La connaissance est extraite des données par des techniques de fouille de données. Ainsi, le système est capable d'évoluer, à mesure que de nouvelles techniques ou solutions gagnent en popularité et deviennent de bonnes pratiques au sein de l'entreprise.

Il a comme objectif de pouvoir recommander aussi bien des approches générales, que de conseiller des paramètres précis à un programmeur CN. Ainsi, il permet une complétion rapide et systématique des tâches routinières grâce à une approche d'apprentissage automatique, mais aussi

une aide à la décision, grâce à des techniques de récupération de cas proches. Ce système est un support au programmeur, dynamique et capable d'évoluer, pour répondre au besoin de l'utilisateur, grâce à des règles qui ne sont pas fixes mais basées sur l'expérience de l'entreprise. Le système proposé a vocation à être utilisé en parallèle du système de FAO, afin de garantir sa stabilité et son indépendance du ou des logiciels et de leurs différentes versions. Il convient de noter que si ce choix apporte en stabilité, le programmeur peut choisir de ne pas se servir du système de propositions. Il sera néanmoins mis à jour, dès que les programmes seront intégrés dans le système d'information d'entreprise. L'architecture globale du système est présentée en Figure 1.

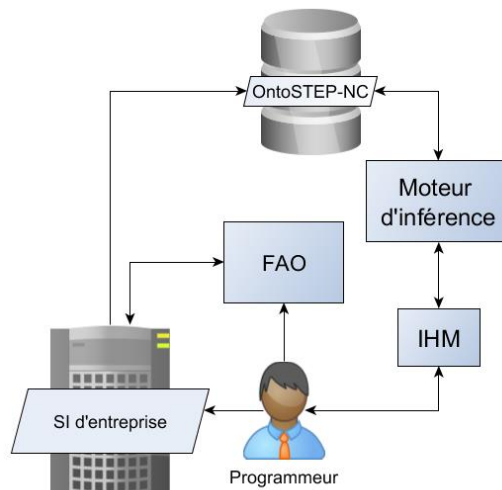


Figure 1. L'architecture globale du système à base de connaissances

Les informations relatives à la planification sont extraites du système d'information d'entreprise, et copiées dans un miroir, OntoSTEP-NC, défini par [DAN 17]. L'information qu'il contient peut ensuite être exploitée pour entraîner un algorithme d'apprentissage automatique, tel qu'introduit en section 3.2.

Les techniques de fouille de données sont considérées pour extraire la connaissance portée par les données d'industrialisation. Il est possible de les appliquer dans le cadre proposé, car elles sont conservées sous un format unique, dans la base ontologique. Il est possible de les extraire sous un format, tabulaire, directement exploitable par les techniques usuelles de fouille de données.

L'utilisateur final est à l'initiative des tâches sur lesquelles déployer l'aide à la décision, et peut proposer de nouveaux problèmes à considérer.

Une fois ces techniques mises en place, l'utilisateur sélectionne une des tâches prédéfinies, et renseigne les paramètres conditionnant la prise de décision dans le cas considéré. L'outil de fouille de données est ainsi capable de proposer une solution – ou plusieurs, avec leurs scores de confiance associés – que le programmeur peut appliquer immédiatement.

Le format considéré permet donc de mettre en place et d'utiliser facilement des techniques de fouille pour assister la prise de décision sur des problèmes de programmation de CN. La mise en place de ces techniques est décrite dans la prochaine partie de cette section.

3.2. Moteur d'inférence : aide au choix de la technique de fouille de données

La revue de [GE 17] souligne le nombre important de techniques liées au domaine de la fouille de données, ainsi que l'importance de choisir une technique en fonction du problème considéré. Il convient donc d'abord d'établir une liste des techniques à considérer [HAN 11]. Dans le cadre de

cette proposition, ne sont considérées que les techniques d'analyse statistique, ainsi que l'apprentissage automatique supervisé. [KOT 07] a passé en revue les principales techniques d'apprentissage automatique supervisé avec comme objectif la classification.

Certaines des techniques considérées sont présentées en Tableau 1. Cette liste est amenée à évoluer, et les différentes techniques et méthodes présentées peuvent présenter des variantes, ou être combinées pour obtenir des techniques hybrides.

Techniques de fouille de données	
Arbre de décision	Réseau de neurones
Machine à vecteur support	Réseau bayésien
Régression	Ensembles approximatifs
k-moyenne	k-proches voisins
Analyse / régression en composantes principales	

Tableau 1. Techniques considérées pour la fouille de données

Il convient ensuite d'établir une liste de critères pour classier les différentes techniques, et évaluer leur capacité à traiter un problème donné.

Une liste non exhaustive de critères est présentée en Tableau 2. Chaque application a des besoins différents. La première étape pour automatiser cette application est d'établir ces besoins, et de les comparer aux performances et capacités des techniques sur les critères définis ci-dessus.

Critères
Type d'objectif (classification, régression, génération...)
Intelligibilité du processus de décision
Type de paramètres manipulés
Type de sortie
Gestion des données inexactes
Quantité de données nécessaires à la technique
Capacité à étendre le domaine d'étude (extensibilité)

Tableau 2. Critères de choix d'une technique de fouille de données

Une fois la liste des techniques capables établie, il convient de suivre les étapes décrites en Figure 2. Les techniques apprenantes nécessitent d'être entraînées, et une étape d'évaluation du modèle, telle que définie par [RAS 18] permet de choisir entre différentes techniques capables. Il est aussi possible, lorsque plusieurs techniques sont capables, d'intégrer leurs résultats dans une structure hybride [VAL 02].

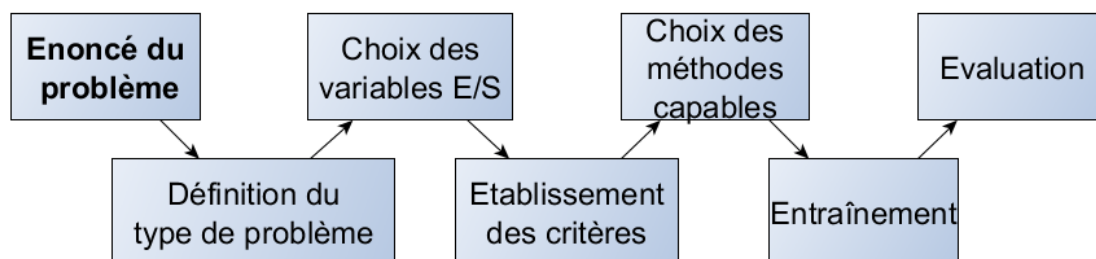


Figure 2. Méthodologie d'application de techniques de fouille

La prochaine partie traitera d'un exemple d'enchaînement de tâches élémentaires, et du choix de techniques à appliquer pour ces tâches.

3.3. Exemple d'utilisation

Afin d'usiner une entité, le programmeur doit prendre plusieurs décisions ;

- 1—Quelle séquence d'opérations réaliser ?
- 2—Quel outil utiliser dans chaque opération ?
- 3—Quels paramètres de coupe adopter pour chaque opération ?

Ces questions ont chacune un objectif différent. Ainsi, les critères de choix d'une technique diffèrent. Le Tableau 3 résume les critères de choix pour ces trois questions.

Critère	1 –	2 –	3 –
Type d'objectif	Génération de séquence	Classification ou régression	Classification ou régression
Attributs	Continus et discrets	Continus et discrets	Discrets
Type de sortie	Séquence d'étiquettes	Continue et discrète	Continue
Extensibilité	Non	Oui ou non	Oui

Tableau 3. Critères limitants appliqués aux trois tâches considérées

La première tâche consiste à générer la séquence d'opérations à effectuer, pour obtenir une entité donnée. Les paramètres d'entrée sont le type d'entité, ainsi que ses dimensions. La sortie est une séquence de types d'opérations. Le modèle n'est pas amené à prendre en considération de nouvelles entrées ou sorties. D'après ces critères, les réseaux de neurones récurrents sont à même de générer une séquence à partir d'un nombre fixe de paramètres d'entrée [VIN 15].

La seconde tâche consiste à sélectionner un outil pour chaque opération. Elle peut être pensée comme une classification (associer à la donnée d'une opération et des dimensions de l'entité une référence d'outil). Considérer cette opération comme une étape de classification rend obligatoire le critère d'extensibilité, car de nouveaux outils sont régulièrement disponibles à la vente. Une autre approche est de considérer ce problème comme composite, mêlant classification (produire un type d'outil) et régression (proposer des dimensions capables). Cette approche n'a plus à satisfaire le

critère d'extensibilité. Plusieurs techniques satisfont ainsi les critères de performance de cette tâche, dont les réseaux de neurones artificiels ou les machines à vecteur support. Les critères absents du Tableau 2 servent ainsi à départager les différentes techniques capables.

Enfin, le troisième objectif, de choix des paramètres de coupe, nécessite d'associer, à une référence outil et à une référence matériau, un nombre fixé de paramètres de coupe. Il y a un fort besoin d'extensibilité, pour prendre en compte de nouveaux outils ou matériaux. Ainsi, une analyse statistique des paramètres de coupe les plus fréquemment utilisés est la technique la plus adaptée à ce besoin.

4. Etude de cas : usiner une poche fermée rectangulaire

L'étude de cas présenté ici s'applique à l'opération et sa stratégie associée (telle que définie par l'élément "two5D_milling_strategy" de la norme ISO14649), en fonction d'une entité d'usinage, de ses dimensions caractéristiques et du matériau. La première partie de cette section se concentre sur la définition de la tâche et l'application d'une technique compétente pour la tâche.

4.1. Prétraitement des données

Les entités choisies sont des poches rectangulaires. Les paramètres sélectionnés en entrée sont les dimensions de la poche (longueur, largeur et hauteur), le rayon de la zone de transition entre les côtés (filet) et entre les côtés et le fond (rayon du fond). La poche est soit une poche ouverte, soit une poche fermée (ouverte, booléenne). Les sorties requises sont le type d'opération (opération), la stratégie (stratégie) et l'épaisseur de matière à conserver après l'application d'une opération donnée (tolérances, fond et côtés).

Le jeu de données utilisé pour la formation des techniques d'apprentissage est construit comme suit :

1. Une première partie du jeu de données est extraite des programmes CN des pièces des partenaires industriels du consortium du projet (voir remerciements). Cinq projets, totalisant huit opérations, ont été utilisés. Ces données représentent un programme qui vise à produire des pièces réelles, qui a été simulé, validé, joué et qui a donné des résultats satisfaisants.

2. Une deuxième partie du jeu de données a été construite avec l'aide des programmeurs CN du consortium. Les pièces (comme dans la Figure 3) ont été fabriquées pour les besoins de cette expérimentation. Le programme, d'autre part, est supposé valide, et provenant de l'expertise de programmeurs professionnels. Au total, 32 opérations sont créées de cette façon.

3. Une troisième partie du jeu de données a été créée artificiellement, suivant des règles d'experts et de bonnes pratiques communes, sur des pièces personnalisées. Cela représente 69 opérations.

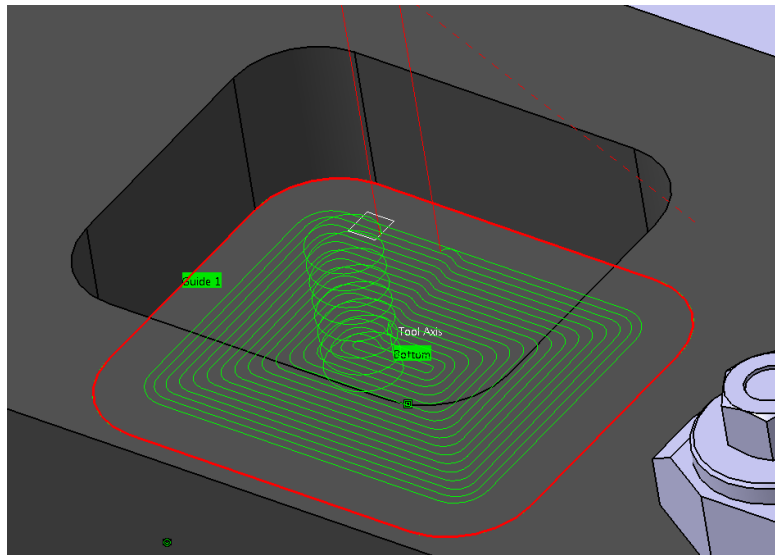


Figure 3. *Fonctionnalité sur mesure pour la constitution d'un jeu de données d'usinage de poches*

La poche rectangulaire est à la fois un problème de classification (type d'opération, stratégie) et de régression (tolérances).

La technique choisie est le réseau neuronal, pour sa capacité à effectuer à la fois la classification et la régression en même temps, et ses formats d'entrée et de sortie flexibles. Son interopérabilité est faible et, selon sa structure, elle peut nécessiter une grande quantité de données. Il est à noter que les sous-tâches peuvent être réalisées par d'autres techniques - l'opération, la stratégie et la finition peuvent être considérées comme trois tâches de classification différentes.

4.2. Réseau de neurones : conception et apprentissage

L'architecture choisie est un réseau neuronal à quatre couches, divisé en une couche "entrée", deux couches "cachées" et une couche "sortie".

Le réseau est conçu pour fournir une combinaison unique de type d'opération, de stratégie et de tolérances à chaque "état" de la poche à usiner. Une fois que l'opération, la stratégie et les allocations d'outils sont connus, le nouvel "état actuel" est mis à jour et une deuxième exécution du réseau donne l'opération suivante à appliquer. Ce cycle est répété jusqu'à ce que la poche soit usinée conformément à ses spécifications.

Le réseau est divisé en trois branches différentes. La première branche se concentre sur la détermination du type et de la qualité de l'opération, la deuxième sur la stratégie à adopter, et la troisième sur les tolérances et conditions de l'outil à utiliser.

La couche d'entrée est commune aux trois réseaux. Les couches cachées et de sortie sont différentes pour chaque branche.

La couche d'entrée est composée de 15 perceptrons - neurones sans fonction d'activation - ("L", "l", "h", "Rcorner", "Rbot", "Open", "Al", "Ti", "Steel", "isPlanar", "L_act", "w_act", "h_act", "Rcorner_act", "Rbot_act") représentant le logement cible et l'état actuel de la pièce à usiner. Le paramètre "L" est la plus grande des deux dimensions du contour de la poche.

Prenons l'exemple d'une poche rectangulaire. Ses dimensions sont décrites (50 mm x 50 mm x 25 mm, avec un filet de 5 mm), son matériau est donné - la poche donnée doit être usinée à partir d'une pièce brute en aluminium - et son état actuel est décrit comme une face plane. Après

l'application d'une opération, la description évolue, et la partie "état actuel" de l'entrée est modifiée. Après l'ébauche de la pièce, son "état actuel" montre une poche de 44mm x 44mm x 22mm, avec un filet de 15mm. Après l'opération de finition latérale, la poche a la longueur et la largeur requises (50mm x 50mm).

Pour normaliser l'entrée, on divise les dimensions par la longueur de la poche à usiner, puis on divise la longueur de la poche par un nombre fixe, en fonction du jeu de données (ici, 150 mm a été choisi comme longueur maximale d'une poche), ce qui donne une description de la poche dont les valeurs sont comprises entre 0 et 1.

Deux couches cachées sont utilisées, pour permettre l'abstraction des données d'entrée. Chaque neurone individuel de ces couches est relié à chacun des neurones des couches précédentes, formant une couche dense. Les neurones des couches cachées utilisent la fonction d'activation tanh. Cette fonction permet des sorties normalisées (entre -1 et 1), et permet une rétropropagation efficace de la perte, grâce à son gradient élevé autour de 0. Tanh permet une convergence plus rapide des réseaux neuronaux, grâce à son gradient élevé, et une sortie limitée.

Plusieurs tests ont été effectués pour sélectionner le nombre de neurones dans chaque couche cachée et la configuration la plus efficace a été de 15 neurones dans chaque couche.

La couche de sortie se compose d'un ensemble de neurones avec des fonctions d'activation spécifiques, en fonction du but du réseau.

La couche de sélection des opérations utilise 4 + 1 neurones. Les 4 premiers neurones codent l'opération à appliquer ("bot_side_mill", "plane_mill", "side_mill", "side_mill", "drilling") et le neurone final caractérise la qualité de l'opération - opération de finition ou non. L'opération utilise une valeur de 1 pour l'opération sélectionnée et de 0 pour toutes les autres - et la qualité est encodée en booléen. La fonction d'activation de tanh est utilisée pour le neurone de qualité, afin d'avoir une forte délimitation entre les opérations de finition et les opérations de non finition. Les fonctions de perte associées à ce réseau sont divisées en deux. La première fonction de perte caractérise la performance de la sélection de la stratégie, tandis que la seconde se concentre sur le neurone de qualité. La première fonction de perte, est la fonction catégorique_crossentropy, qui favorise les sorties fortement séparées. La deuxième fonction de perte est la binary_crossentropy, qui permet d'évaluer une seule valeur booléenne.

Le deuxième réseau se concentre sur le choix de la stratégie. Il est constitué de sept neurones ("Unidir_mill", "Bidir_mill", "Contour_para", "Bidir_contour", "Contour_spiral", "Center_mill" et "Explicit"), décrivant la stratégie. La fonction d'activation associée est softmax, et la perte est l'entropie croisée catégorique, comme pour la sélection de l'opération.

Le troisième réseau fournit deux variables, les tolérances de l'opération sélectionnée. Il est constitué de quatre neurones ("Allow_bot", "Allow_side", "Rcoin_rea" et "Rbot_rea") et utilise la fonction d'activation max(0,x). Il délivre une valeur positive ou nulle, adaptée aux surépaisseurs d'une opération d'usinage. La fonction de perte est définie comme l'erreur quadratique moyenne de la prédiction.

Les paramètres adoptés pour l'évaluation de ces réseaux sont les fonctions de perte de chaque réseau.

Le jeu de données est divisé au hasard en deux ensembles. Le premier ensemble comprend 80 % de l'ensemble des données et sert à la formation des réseaux. Le second représente 20% de l'ensemble des données et sert à la validation des réseaux. L'étape de validation consiste à exécuter les réseaux sur des cas inédits et à calculer leur exactitude à l'aide des mesures susmentionnées.

4.3. Post traitement et interprétation des résultats

En fonction de la valeur d'initialisation des paramètres du réseau et de la sélection des données de validation, certains entraînements donnent de meilleurs résultats que d'autres. Les résultats illustrés à la figure 4 sont les meilleurs de 20 passages, en utilisant les paramètres décrits au point 4.2. Néanmoins, tous les cycles d'entraînement ont entraîné une stagnation ou une diminution des pertes au cours de la période d'entraînement, ce qui indique que le modèle est capable d'apprendre à partir des données qui lui sont fournies.

Le jeu de données utilisé pour la formation de ces réseaux est petit par rapport aux normes de la littérature - moins de 100 opérations, comparativement à des milliers dans la plupart des publications d'apprentissage automatique - ce qui explique la variabilité des pertes entre deux opérations.

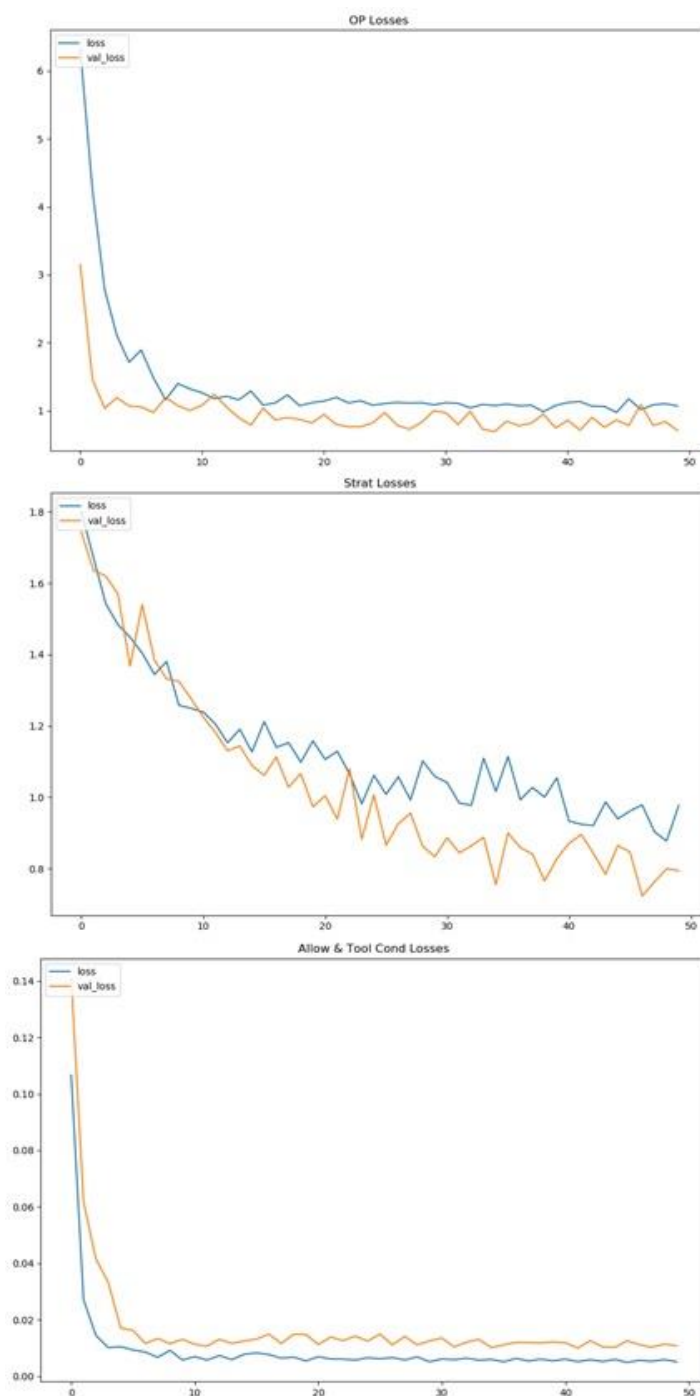


Figure 4. Évaluation des performances des trois réseaux pendant la formation

Les valeurs numériques de la performance de ces modèles peuvent être obtenues à l'aide de la formule d'entropie croisée. Il est défini, sur un problème de classification de classe M, comme :

$$-\sum_{c=1}^M y_{o,c} \log(p_{o,c}) \quad [1]$$

Avec :

$y_{o,c}$ l'indicateur binaire si l'étiquette de classe c est la classification correcte pour l'observation o

$p_{o,c}$ la probabilité prévue d'observation o étant de classe c

Ainsi, pour un problème à quatre classes, la perte associée à une distribution aléatoire est égale à 1,38. Avec une perte moyenne de 0,53 sur la sélection des opérations au cours des 10 derniers entraînements, ce réseau a une précision de 58,8%, 30 points supérieurs à une distribution aléatoire (25% de précision). Sa qualité en tant que prédicteur pourrait être améliorée en augmentant encore la quantité et l'homogénéité des données d'apprentissage. Le jeu de données a été construit avec des données provenant de différentes sources, donc de différents programmeurs. Chaque programmeur a ses habitudes concernant les opérations à utiliser, de sorte que le réseau est formé sur différentes expertises. Pour un même cas, le jeu de données peut présenter différentes séquences d'opérations, avec le même résultat. La perte moyenne sur les 10 derniers entraînements du réseau pour les stratégies est de 0,82, soit une précision de 44%, 30 points de plus que la distribution aléatoire (précision de 14,2%).

5. Conclusion

Cet article propose un cadre KBE pour assister un programmeur CN pendant le processus de FAO. Il présente l'intégration des données d'industrialisation au sein d'une base ontologique basée sur la suite de normes STEP, afin d'indexer les données relatives à l'activité d'usinage. Les informations sont extraites du système d'information de l'entreprise puis sont traduites dans un format basé sur STEP-NC, pour éliminer les problèmes d'hétérogénéité de format. Une méthode de mise en œuvre d'une application personnalisée des techniques de fouilles de données est proposée, en cinq étapes. Les étapes sont la définition de l'outil nécessaire, l'établissement de critères et la comparaison avec un classement des techniques en fonction de ces critères, la formation de la ou des techniques capables, la comparaison et éventuellement le renforcement, et enfin, une étape de validation industrielle. L'étude de cas suit les étapes susmentionnées pour créer un outil d'aide à la décision visant à recommander une séquence d'opérations à effectuer pour usiner une poche rectangulaire, obtenant des résultats acceptables sur un petit ensemble d'entraînement. L'approche proposée et démontrée dans cet article donne des résultats encourageants. L'étape suivante pour la validation d'une telle approche est la population à grande échelle de la base de cas, afin de permettre de tester davantage l'outil de recommandation.

Remerciements

Ce travail a été réalisé dans le cadre du projet FUI 21 LUCID. Les auteurs souhaitent remercier tous les partenaires du projet pour leurs apports et remarques durant l'élaboration et la présentation des idées et concepts développés ici.

Bibliographie

[ALT 89] ALTING, L. AND ZHANG, H., «Computer Aided Process Planning: the state-of-the-art survey», *International Journal of Production Research*, n° 27(4), p. 553–585, 1989.

- [ARE 00] AREZOO, B., RIDGWAY, K. AND AL-AHMARI, A. M. A., «Selection of cutting tools and conditions of machining operations using an expert system», *Computers in Industry*, n°42(1), p. 43–58, 2000.
- [BER 08] BERMELL-GARCIA, P. AND FAN, I. S., «Practitioner requirements for integrated Knowledge-Based Engineering in Product Lifecycle Management», *International Journal of Product Lifecycle Management*, n°3(1), p. 3–20, 2008.
- [CAO 13] CAO, Y., CAO, S., BAI, Y. AND YANG, X. F., «CBR-Based Cutting Parameter Selection System and Searching Algorithm for Metal Machining Operations», *Advanced Materials Research*, n°710, p. 554–557, 2013.
- [DAN 17] DANJOU, C., LE DUIGOU, J. AND EYNARD, B., «Closed-loop manufacturing process based on STEP-NC», *International Journal on Interactive Design and Manufacturing*, n°11(2), p. 233–245, 2017.
- [DEN 16a] DENKENA, B., DITTRICH, M. A. AND UHLICH, F., «Augmenting Milling Process Data for Shape Error Prediction», *Procedia CIRP*, n°57, p. 487–491, 2016a.
- [DEN 16b] DENKENA, B., DITTRICH, M. A. AND UHLICH, F., «Self-optimizing cutting process using learning process models», *Procedia Technology*, p. 221–226, 2016b.
- [DEN 14] DENKENA, B., SCHMIDT, J. AND KRÜGER, M., «Data Mining Approach for Knowledge-based Process Planning», *Procedia Technology*, p. 406–415, 2014.
- [FAN 08] FAN, I.-S. AND BERMELL-GARCIA, P., «International Standard Development for Knowledge Based Engineering Services for Product Lifecycle Management», *Concurrent Engineering*, n°16(4), p. 271–277, 2008.
- [FAY 96] FAYYAD, U., PIATETSKY-SHAPIO, G. AND SMYTH, P., «From Data Mining to Knowledge Discovery in Databases», *AI Magazine*, n°17(3), p. 37–54, 1996.
- [GE 17] GE, Z., SONG, Z., DING, S. X. AND HUANG, B., «Data Mining and Analytics in the Process Industry: the Role of Machine Learning», *IEEE Access*, n°3536(c), p. 1–27, 2017.
- [HAN 16] HAN, J.H. AND CHI, S.Y., «Consideration of manufacturing data to apply machine learning methods for predictive manufacturing», *International Conference on Ubiquitous and Future Networks*, p. 109–113, 2016.
- [HAN 11] HAN, J., PEI, J. AND KAMBER, M., *Data Mining: Concepts and Techniques*, Edited by Elsevier, 2011.
- [HUE 11] HUET, G., FORTIN, C., MCSORLEY, G. AND TOCHE, B., «The Management Of Manufacturing Processes Using Complementary Information Structures», *Proceedings of the 18th International Conference on Engineering Design*, n°6, p. 183–191, 2011.
- [ISO 03] ISO14649-1, *STEP NC, ISO 14649-1, Part 1: Overview and fundamental principles*, 2003.
- [JOS 08] JOSHI, R. S., KUMAR, N. AND SHARMA, A., «Setup Planning and Operation Sequencing Using Neural Network and Genetic Algorithm», *Fifth International Conference on Information Technology: New Generations*, p. 396–401, 2008.
- [KOT 07] KOTSIANTIS, S. B., «Supervised machine learning: A review of classification techniques», *Informatica*, n°31, p. 249–268, 2007.
- [LU 90] LU, S.C.Y., «Machine learning approaches to knowledge synthesis and integration tasks for advanced engineering automation», *Computers in Industry*, n°15(1–2), p. 105–120, 1990.
- [MAN 16] MANGAL, A. AND KUMAR, N., «Using big data to enhance the bosch production line performance: A Kaggle challenge», *Proceedings - IEEE International Conference on Big Data*, p. 2029–2035, 2016.
- [PER 14] PERZYK, M., KOCHANSKI, A., KOZŁOWSKI, J., SOROCZYNSKI, A. AND BIERNACKI, R., «Comparison of data mining tools for significance analysis of process parameters in applications to process fault diagnosis», *Information Sciences*, n°259, p. 380–392, 2014.
- [RAS 18] RASCHKA, S., «Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning», arXiv:1811.12808, 2018.
- [LAR 11] LA ROCCA, G., *Knowledge Based Engineering Techniques to Support Aircraft Design and Optimization*, Thèse de doctorat, Technische Universiteit Delft, 2011.
- [ROJ 17] ROJEK, I., «Technological process planning by the use of neural networks», *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, n°31(1), p. 1–15, 2017.

- [TOR 13] TORRES-TREVIÑO, L. M., ESCAMILLA-SALAZAR, I. G., GONZÁLEZ-ORTÍZ, B. AND PRAGA-ALEJO, R., «An expert system for setting parameters in machining processes», *Expert Systems with Applications*, n°40(17), p. 6877–6884, 2013.
- [VAL 02] VALENTINI, G. AND MASULLI, F., «Ensembles of Learning Machines», *Lecture Notes in Computer Science*, 2002.
- [VIN 15] VINYALS, O., TOSHEV, A., BENGIO, S. AND ERHAN, D., «Show and tell: A neural image caption generator», *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, p. 3156–3164, 2015.
- [WUE 16] WUEST, T., WEIMER, D., IRGENS, C. AND THOBEN, K.-D., «Machine learning in manufacturing: advantages, challenges, and applications», *Production & Manufacturing Research*, n°4(1), p. 23–45, 2016.