

Exploitation des données massives au pilotage multi-objectifs de l'Industrie 4.0

Amel SOUIFI

CY Paris Cergy Université et Le Cnam

amel.souifi@lecnam.net

November 26, 2021

Contexte et problématique
État de l'art
Cas d'étude: modèle de simulation
Quantité de données
Qualité de données
Cas d'application
Conclusion et perspectives
Références

Plan

- 1 Contexte et problématique
- 2 État de l'art
- 3 Cas d'étude: modèle de simulation
- 4 Quantité de données
- 5 Qualité de données
- 6 Cas d'application
- 7 Conclusion et perspectives
- 8 Références

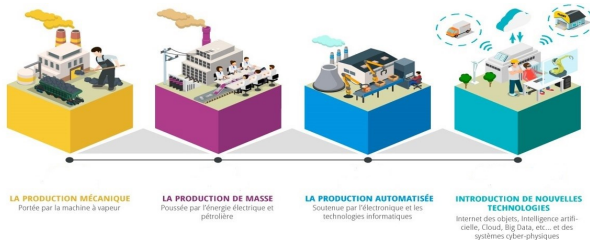
Contexte et problématique
État de l'art
Cas d'étude: modèle de simulation
Quantité de données
Qualité de données
Cas d'application
Conclusion et perspectives
Références

Plan

- 1 Contexte et problématique
- 2 État de l'art
- 3 Cas d'étude: modèle de simulation
- 4 Quantité de données
- 5 Qualité de données
- 6 Cas d'application
- 7 Conclusion et perspectives
- 8 Références

Contexte et problématique
État de l'art
Cas d'étude: modèle de simulation
Quantité de données
Qualité de données
Cas d'application
Conclusion et perspectives
Références

Contexte



Apports: satisfaction clients, flexibilité, productivité, satisfaction des employés, etc.

Défis: Présence de grandes quantités de données, automatisation et réduction de la place de l'homme

Problématique

- Définition de méthodes pour déterminer la quantité suffisante de données.
- Étudier l'influence de la qualité de données sur la prise de décision.
- Proposer un méthode d'aide à la décision tenant compte des données massives.

Contexte et problématique
État de l'art
Cas d'étude: modèle de simulation
Quantité de données
Qualité de données
Cas d'application
Conclusion et perspectives
Références

Plan

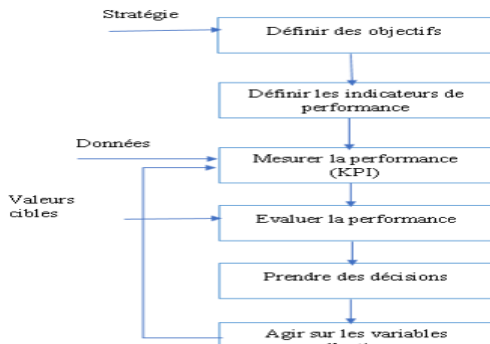
- 1 Contexte et problématique
- 2 État de l'art
- 3 Cas d'étude: modèle de simulation
- 4 Quantité de données
- 5 Qualité de données
- 6 Cas d'application
- 7 Conclusion et perspectives
- 8 Références

Performance

- Définition 1 : l'atteinte d'objectifs prédéfinis par un individu, un groupe ou une organisation.
- Définition 2 : la performance organisationnelle peut être également définie comme un rapport entre une production de valeur et une consommation de ressources.

Pilotage de performance

- Décider dynamiquement des consignes pertinentes à donner à un système soumis à perturbation pour atteindre un objectif donné décrit en termes de maîtrise de performances” .



- Contexte et problématique
- État de l'art
- Cas d'étude: modèle de simulation
 - Quantité de données
 - Qualité de données
 - Cas d'application
- Conclusion et perspectives
- Références

Indicateurs de performance

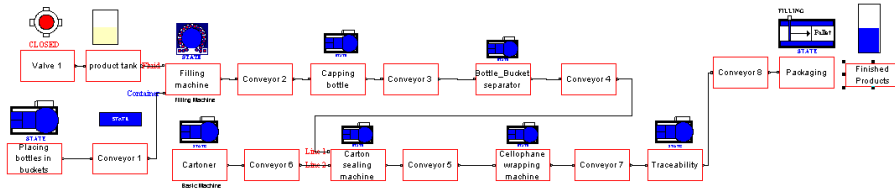
Selon la norme ISO 22400, un indicateur de performance est le niveau quantifiable de réalisation d'un objectif critique.

Contexte et problématique
État de l'art
Cas d'étude: **modèle de simulation**
Quantité de données
Qualité de données
Cas d'application
Conclusion et perspectives
Références

Plan

- 1 Contexte et problématique
- 2 État de l'art
- 3 Cas d'étude: modèle de simulation**
- 4 Quantité de données
- 5 Qualité de données
- 6 Cas d'application
- 7 Conclusion et perspectives
- 8 Références

Modèle de simulation



Contexte et problématique
État de l'art
Cas d'étude: modèle de simulation
Quantité de données
Qualité de données
Cas d'application
Conclusion et perspectives
Références

Plan

- 1 Contexte et problématique
- 2 État de l'art
- 3 Cas d'étude: modèle de simulation
- 4 Quantité de données**
- 5 Qualité de données
- 6 Cas d'application
- 7 Conclusion et perspectives
- 8 Références

Conditions de pertinence de données

Useful KPI	Useful data
Aligné	Des données alignées à la stratégie de l'entreprise
Correct	
Précis	Données de qualité
D'actualité	
Peu coûteux	Données de coût acceptable
Complet	Une quantité suffisante de données
Pertinent	Données pertinentes

Quantité d'information dans les données

La quantité d'information peut être mesurée grâce à l'entropie de Shannon

Pour une variable aléatoire:

$$H(X) = - \sum_{x \in X} p(x) \log(p(x))$$

Pour un ensemble de variables aléatoires:

$$\begin{aligned} H(X_1, X_2, \dots, X_n) &= \sum_{i=1}^n H(X_i / X_{i-1}, \dots, X_1) \\ &= - \sum_{i=1}^n \sum_{x_1, \dots, x_i} p(x_1, \dots, x_i) \log(p(x_i / x_{i-1}, \dots, x_1)) \end{aligned}$$

Plus de données signifie-t-il plus d'information?

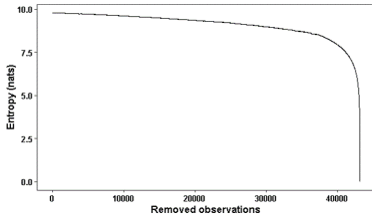


Figure: Variation de l'entropie en fonction du nombre d'observations éliminées

Aucune observation éliminées-----
Entropie=9,79 nats
30 000 observations éliminées-----
Entropie = 8,99 nats (8% de perte)
40 000 observations éliminées-----
Entropie = 7,95 nats (19% de perte)

43000 observations éliminées-----
Entropie = 5,26 nats (46% de perte)

Figure: Interprétation des résultats

Plus de données signifie-t-il plus d'information?

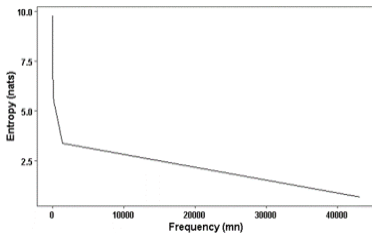


Figure: Variation de l'entropie en fonction de la fréquence de collecte de données

Collecte de données chaque minute-----
Entropie=9,79 nats
Collecte de données chaque 3mn-----
Entropie =9,06 nats (7% de perte)
Collecte de données chaque 10 mn---
Entropie = 8,24 nats (15%) de perte
Collecte de données chaque 3 heures---
Entropie = 5,46 nats (44% de perte)

Figure: Interprétation des résultats

Plus de données signifie-t-il plus d'information?

Conclusion 1

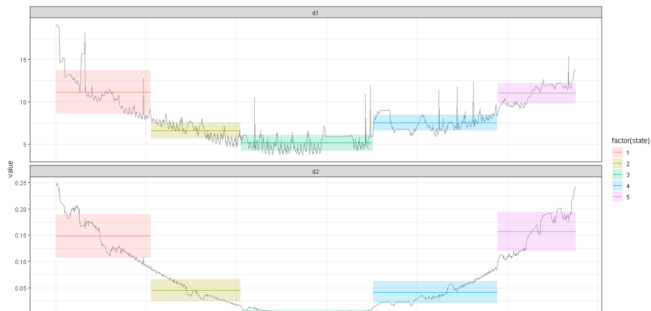
Plus de données ne fournit pas obligatoirement plus d'information

Conclusion 2

La fréquence de collecte de données doit être variable en fonction du temps afin de répondre à la dynamique du système et tenir compte des perturbations.

Comment définir les observations nécessaires et suffisantes

La définition d'une fréquence de collecte de données peut être formulée comme étant un problème de segmentation de séries temporelles visant la détection des points de changement dans les observations.



Choix d'indicateurs de performance

- Choisir quelles dimensions de performance afin de mesurer la performance globale.
- Choisir quels indicateurs pour ne pas surcharger le tableau de bord.
- Hiérarchiser les indicateurs (permet de choisir par exemple quels indicateurs afficher en premier et quels autres laisser sur la demande de l'opérateur.
→ Analyse en composantes principales, clustering de variables

Plan

- 1 Contexte et problématique
- 2 État de l'art
- 3 Cas d'étude: modèle de simulation
- 4 Quantité de données
- 5 Qualité de données**
- 6 Cas d'application
- 7 Conclusion et perspectives
- 8 Références

Incertitude de données

- Un indicateur de performance est représenté par: (KPI, m) , KPI représente un intervalle de tolérance contenant la valeur de l'indicateur et m une mesure de confiance.
- La détermination de m est basée sur la théorie de croyance

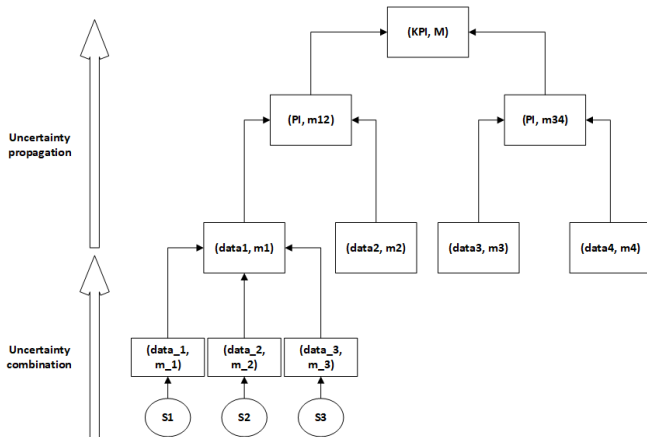
Pourquoi la théorie de croyance?

- Une modélisation relativement simple de l'incertitude.
- Modélise à la fois l'imprécision et l'incertitude de données.
- Offre un cadre important pour la fusion de données.

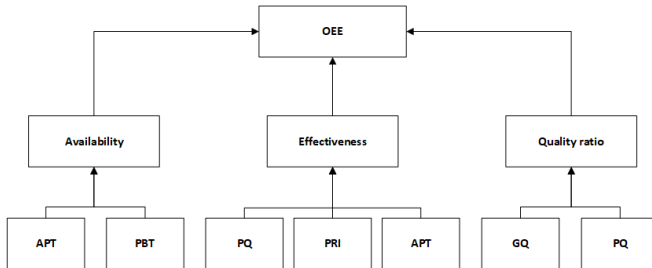
Éléments de base de la théorie de croyance

- La fonction de masse exprime un degré de confiance dans un élément A de $P(\Omega)$.
- La masse attribuée à $A = \{H1, H2\}$ par exemple est attribuée strictement à A et ne peut être allouée à aucune hypothèse plus restrictive.
- Une masse attribuée à une hypothèse simple exprime le degré de confiance de cette hypothèse par rapport aux autres.
- Une masse attribuée à une hypothèse composée exprime la confusion associée au manque d'information pour décider entre une hypothèse ou une autre

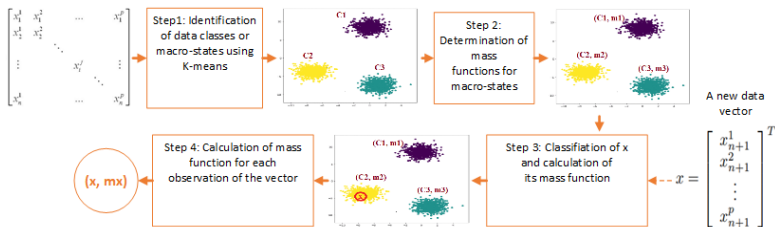
Méthodologie



Hierarchisation des indicateurs de performance



Démarche de calcul de la fonction de masse basée sur l'historique de données



Détermination de la fonction de masse basée sur la distance

On calcule de la confiance dans la valeur de x en se basant sur la distance le séparant de ses proches voisins: ici k est pris égal à 2 ou 3.

$$m_q = 1 - \prod_{x_i \in \Phi^q} (1 - \alpha_0 \phi(d^i))$$

avec: $\alpha_0 = 0.95$

$\phi(d^i) = \exp(-d)$

La fonction de masse d'un nouveau vecteur de données est égale au produit de la fonction de masse du groupe auquel il appartient et la fonction de masse basée sur la distance.

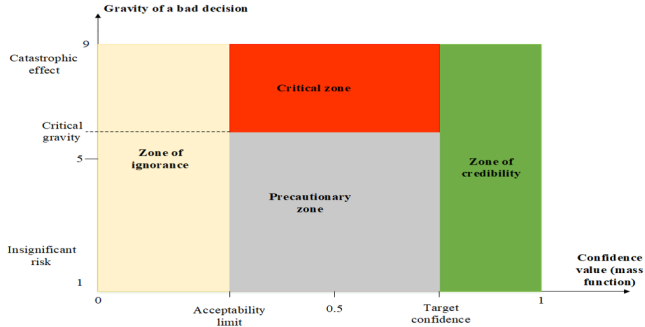
Propagation d'incertitude aux indicateurs de performance

Indicateur = $F(\text{donnée } 1, \text{donnée } 2, \dots, \text{donnée } n)$

On propage l'incertitude liée aux données aux indicateurs via **le produit cartésien**.

Contexte et problématique
État de l'art
Cas d'étude: modèle de simulation
Quantité de données
Qualité de données
Cas d'application
Conclusion et perspectives
Références

Prise de décision

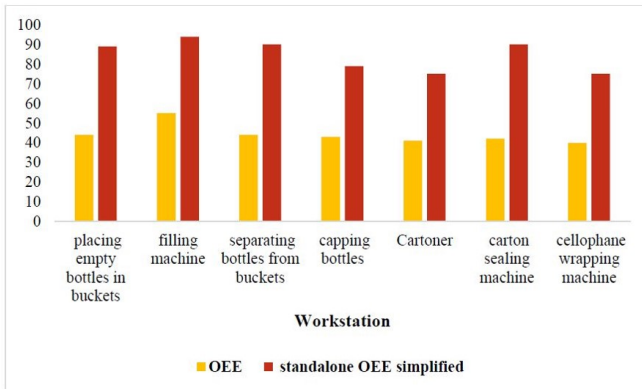


Contexte et problématique
État de l'art
Cas d'étude: modèle de simulation
Quantité de données
Qualité de données
Cas d'application
Conclusion et perspectives
Références

Plan

- 1 Contexte et problématique
- 2 État de l'art
- 3 Cas d'étude: modèle de simulation
- 4 Quantité de données
- 5 Qualité de données
- 6 Cas d'application**
- 7 Conclusion et perspectives
- 8 Références

Choix de la machine goulot



Calcul de TRS

- **TRS = Taux de disponibilité * taux de qualité * taux de performance**
- On suppose que le taux de qualité est **constant** et est de 0.98.
- **Taux de disponibilité** = temps de production à valeur ajoutée / (Temps de fonctionnement planifié- temps d'arrêt planifié)
- **Taux de performance** = (temps planifié par unité * Nombre de pièces produites) / temps de production à valeur ajoutée = Nombre de pièce produites / (cadence nominale * temps de production à valeur ajoutée)

Données élémentaires nécessaires au calcul de TRS et de ses composantes

- Pour la disponibilité:
 - Variable:** temps de production à valeur ajoutée
 - Constantes:** Temps de fonctionnement planifié et temps d'arrêts planifiés
- Pour la performance:
 - Variables:** Nombre de pièces produites et temps de production à valeur ajoutée
 - Constante:** cadence nominale

Données élémentaires nécessaires au calcul de TRS et de ses composantes

P

our les constantes on suppose qu'on est certain de leur valeur.

O

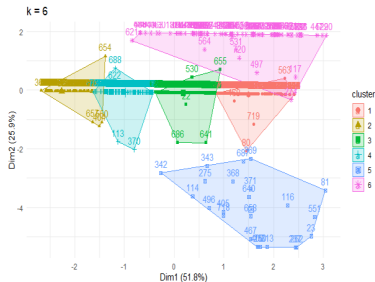
n s'intéresse seulement à l'incertitude associée aux variables à savoir: Le temps de production à valeur ajoutée Le nombre de pièces produites

Définition de la matrice de données

Matrice contenant: le nombre de pièces produites par heure, le temps de production à valeur ajoutée chaque heure, temps de panne et temps d'arrêts planifiés.

La simulation est faite sur un mois et les données sont collectées chaque heure. On obtient une matrice de données de taille: $719^* 4$

Définition des différents macro-états du système



1. Premier groupe : Il correspond à une sous-alimentation de la machine pendant 65% du temps. Cette information est confirmée par les données de simulation concernant le temps de sous-alimentation de la machine par heure.
2. Deuxième groupe : Ce groupe correspond au fonctionnement nominal de la machine. En effet, le temps de production est en moyenne de 55 minutes par heure et le nombre de pièces produites par heure est de 1489.
3. Troisième groupe : Dans ce groupe, la machine est affamée pendant 50% du temps par heure.
4. Quatrième groupe : Cet ensemble de données correspond à une sous-alimentation de la cellophaneuse pendant 30% du temps.
5. Cinquième groupe : Il correspond à l'état de défaillance du système caractérisé par un temps de défaillance moyen de 30 minutes par heure.
6. Sixième groupe : Ce dernier groupe de données correspond à un état de sous-alimentation et à la présence d'arrêts planifiés.

- Contexte et problématique
- État de l'art
- Cas d'étude: modèle de simulation
- Quantité de données
- Qualité de données
- Cas d'application**
- Conclusion et perspectives
- Références

Nouveaux vecteurs de données

Id	Units produced	APT	Time failed	Time stopped	Class	Mass function
1	808	25.6	0.44	1.27	4	0.99711
2	1192	37.79	0.45	5.16	3	0.16531
3	604	19.2	0.2	1.12	1	0.98767
4	167	5.25	0.12	0.24	1	0.99759

Calcul de TRS et son incertitude

TRS (%)	Fonction de masse
38	0.99711
55	0.16531
28	0.98767
6	0.99759

Plan

- 1 Contexte et problématique
- 2 État de l'art
- 3 Cas d'étude: modèle de simulation
- 4 Quantité de données
- 5 Qualité de données
- 6 Cas d'application
- 7 Conclusion et perspectives**
- 8 Références

Conclusion et perspectives

- Pour piloter la performance d'une entreprise à l'ère de l'industrie 4.0, le défi n'est plus la disponibilité des données mais plutôt la qualité et leur pertinence.
- Une définition d'un modèle décisionnel basé sur des données incertaines devrait compléter ce travail.

Plan

- 1 Contexte et problématique
- 2 État de l'art
- 3 Cas d'étude: modèle de simulation
- 4 Quantité de données
- 5 Qualité de données
- 6 Cas d'application
- 7 Conclusion et perspectives
- 8 Références

Berrah, Lamia (2013). "La quantification de la performance dans les entreprises manufacturières : de la déclaration des objectifs à la définition des systèmes d'indicateurs".

Glenn Shafer. A mathematical theory of evidence, volume 42.
Princeton university press, 1976

Thierry Denoeux. A k-nearest neighbor classification rule based on dempster-shafer theory. In Classic works of the Dempster-Shafer theory of belief functions, pages 737–760. Springer, 2008.

Merci de votre attention