

19 RÉPUBLIQUE FRANÇAISE
INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE
COURBEVOIE

11 N° de publication :
(à n'utiliser que pour les
commandes de reproduction)

3 113 151

21 N° d'enregistrement national : 20 08243

51 Int Cl⁸ : G 06 F 17/10 (2020.12), G 06 F 16/245, A 61 K 39/00

12

DEMANDE DE BREVET D'INVENTION

A1

22 Date de dépôt : 03.08.20.

30 Priorité :

43 Date de mise à la disposition du public de la
demande : 04.02.22 Bulletin 22/05.

56 Liste des documents cités dans le rapport de
recherche préliminaire : *Se reporter à la fin du
présent fascicule*

60 Références à d'autres documents nationaux
apparentés :

Demande(s) d'extension :

71 Demandeur(s) : IPPON INNOVATION Société à res-
ponsabilité limitée — FR.

72 Inventeur(s) : ARCHIMBAUD AURORE, RUIZ Anne
et SOUAL Carole.

73 Titulaire(s) : IPPON INNOVATION Société à respon-
sabilité limitée.

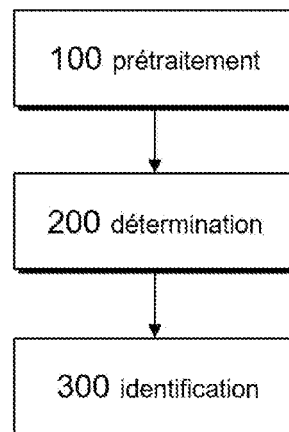
74 Mandataire(s) : IPSIDE.

54 procédé de détermination du niveau d'atypicité d'individus, permettant notamment de détecter statistiquement des individus atypiques dans un cadre multivarié.

57 L'invention concerne un procédé de détermination du niveau d'atypicité d'individus, permettant notamment de détecter statistiquement des individus atypiques dans un jeu de données préalablement récoltées, issues de mesures de paramètres d'individus réalisées par une pluralité de systèmes de mesures, caractérisé en ce qu'il comporte les étapes de:

- prétraitement 100 des données;
- détermination 200 d'un indice d'atypicité multivarié, ledit indice étant transformé par une fonction f pour être compris entre 0 et 1, sur l'ensemble des mesures pour chaque individu, à partir des données prétraitées;
- Identification 300 des individus atypiques.

Figure pour l'abrégié : Fig. 1



FR 3 113 151 - A1



Description

Titre de l'invention : procédé de détermination du niveau d'atypicité d'individus, permettant notamment de détecter statistiquement des individus atypiques dans un cadre multivarié

Domaine technique de l'invention

- [0001] La présente invention appartient au domaine du contrôle de qualité, tel que celui des composants électroniques ou des produits issus de l'industrie pharmaceutique, par exemple des vaccins. Notamment, l'invention concerne un procédé de détermination du niveau d'atypicité d'individus, permettant notamment de détecter statistiquement des individus atypiques dans un cadre multivarié, c'est-à-dire que les individus sont caractérisés par plusieurs tests ou mesures.
- [0002] Dans la présente demande, on désigne par le terme « individu », toute pièce ou lot produit en petite, moyenne ou grande série par des moyens de production industrielle ou autres.
- [0003] Un individu atypique est notamment un individu qui a passé tous les tests avec succès mais qui est statistiquement atypique, et donc, un individu potentiellement non conforme. En effet, l'expérience montre sur les composants électroniques ou les vaccins, que l'atypicité statistique, même sur des pièces ou des lots ayant passés avec succès les tests de contrôle qualité, peut révéler un problème de qualité ou de fiabilité latent.

Technique antérieure

- [0004] La présente invention peut avantageusement être appliquée au secteur de l'industrie des semi-conducteurs. Cette industrie produit des circuits intégrés, dits « composants électroniques », qui sont fabriqués sur des lots de plaquettes de silicium, chaque plaquette comprenant plusieurs centaines de composants électroniques.
- [0005] Afin de s'assurer du fonctionnement de ces composants électroniques, des séries de tests sont effectuées sur chacun des composants alors qu'ils font encore partie d'une plaquette.
- [0006] A chacun de ces tests est associée une ou deux limites de spécification.
- [0007] Les composants électroniques dont la réponse a au moins un test non conforme à la spécification de ce test, c'est-à-dire dont la réponse est en dehors des limites de ladite spécification, sont alors considérés comme défectueux, et sont rejetés lors de leur séparation d'avec la plaquette.
- [0008] Ce contrôle qualité classique peut être complété par des méthodes statistiques de détection d'anomalies, par exemple dans les composants destinés à l'industrie automobile, pour minimiser les problèmes de qualité vus par le client. Il s'agit alors

d'éliminer des pièces bonnes (qui ont passé tous les tests) car elles sont statistiquement anormales et peuvent parfois se révéler mauvaises chez le client. Cet objectif de zéro défaut fait appel à des méthodes dites « univariées » ou « bivariées ».

- [0009] Par exemple, une méthode univariée appelée Part Average Testing (PAT) compare la réponse d'un test d'un composant électronique à la distribution moyenne des réponses de ce test des autres composants électroniques et considère comme composant électronique atypique, un composant électronique dont la réponse est trop distante de la répartition des réponses des autres composants électroniques.
- [0010] Cette méthode n'est généralement pas satisfaisante en pratique car elle mène souvent à de nombreuses fausses alertes puisqu'il suffit qu'une valeur soit à plus de k écarts types de la moyenne sur un seul des p paramètres pour que l'individu soit considéré comme atypique. Plus le nombre p de paramètres augmente, plus il y a de chances pour que l'individu soit atypique sur au moins un test.
- [0011] En utilisant ces méthodes univariées (PAT, etc.) ou bivariées (régressions simples entre un test et un autre et détection des atypiques sur la base de ces régressions) sur un très grand nombre de tests, les fabricants rejettent donc trop de composants, dont de nombreux composants corrects, ce qui les prive de quelques pourcents de leur production, tout en ne leur garantissant cependant pas d'éliminer tous les composants potentiellement défectueux.
- [0012] En effet, ces méthodes sont appliquées sur chaque mesure d'un test de manière indépendante et donc le taux de fausses alertes augmente de façon spectaculaire avec le nombre de mesures. Afin de limiter ce nombre de fausses alertes (et donc le coût par rapport à la détection) les limites statistiques sont choisies de manière souvent trop permissive, au risque de ne pas être suffisantes.
- [0013] Ces méthodes présentent donc le risque de permettre à des composants électroniques comportant un défaut latent qui peut se révéler lors de l'utilisation de la pièce dans le cadre de l'application du client, d'être considérés comme fiables et livrables au client.
- [0014] Cet inconvénient est gênant, d'une part, parce qu'il oblige le fabricant à renvoyer au client une nouvelle pièce de remplacement en cas d'identification du défaut latent, et diminue son niveau de qualité perçu par le client, mais, plus encore, parce que certains de ces composants, bien que d'un coût unitaire faible, sont des composants critiques dans le fonctionnement d'un système plus complexe, par exemple un contrôleur moteur ou un système de freinage ABS. Dans ce cas, une panne du composant peut conduire à un accident grave, dont les conséquences vont très au-delà de la simple valeur financière du composant.
- [0015] Quoique déjà dotées d'une certaine performance, ces méthodes sont donc insuffisantes pour atteindre le zéro défaut.
- [0016] Des solutions ont donc été développées pour répondre à ce besoin avec la mise en

- œuvre de méthodes statistiques multivariées, notamment des méthodes d'analyse en composantes principales ou bien la distance de Mahalanobis et le T^2 de Hotelling.
- [0017] Toutefois, ces solutions ne répondent que partiellement aux attentes du contexte de fabrication industrielle. En effet, comme les méthodes d'analyse en composantes principales ne sont spécialement conçues pour détecter des individus atypiques, elles sont souvent en difficulté pour réaliser une telle tâche.
- [0018] Plus précisément, la distance de Mahalanobis et le T^2 de Hotelling sont des distances multivariées de similarité des individus qui peuvent donc être utilisées pour identifier des individus atypiques. Toutefois, ces méthodes sont en difficulté dès lors que l'atypicité est due seulement à un sous-ensemble de mesures et non pas à la majorité, ce qui est en fait la pratique dans le contexte de fabrication industrielle, au vu des hauts standards de qualité recherchés.
- [0019] D'autres méthodes multivariées existent, telle que la méthode connue par l'homme du métier sous l'acronyme anglophone « LOF », pour « local outlier factor ». Toutefois, cette méthode présente plusieurs inconvénients majeurs : il n'existe pas de règle statistique claire pour identifier les individus atypiques, elle requiert une optimisation d'un paramètre intrinsèque à l'algorithme (le nombre de voisins) et elle demande une capacité de calcul d'autant plus importante que le volume de données augmente ; ce qui représente des contraintes non négligeables. Notamment, compte-tenu de ces inconvénients, l'utilisation en temps réel sur les lignes de production est impossible.
- [0020] Certaines méthodes s'appuient quant à elles sur de la connaissance des défauts à détecter, ce sont alors des méthodes dites « supervisées », ce qui oblige les industriels à calibrer de nouveau la méthode à chaque nouveau produit et/ou ajout de tests dans le processus d'amélioration de la couverture de tests.
- [0021] De nombreuses méthodes se contentent d'attribuer à chaque individu un indicateur d'atypicité sans véritable identification des individus atypiques. Concrètement elles ordonnent ou classent les individus des plus atypiques aux moins atypiques mais ne peuvent assurer que le premier individu soit véritablement atypique.
- [0022] Enfin, de manière générale, la plus grande limitation de ces méthodes réside dans le fait qu'elles ne sont généralement plus applicables dès lors que le nombre de mesures est supérieur au nombre d'individus ou dès qu'il y a de la corrélation entre les variables, ce qui est chose courante dans un contexte industriel ou pharmaceutique. En effet, les données analysées sont souvent issues d'un grand nombre de mesures réalisées par un certain nombre de tests, typiquement des centaines, voire des milliers, sur chaque composant électronique ou individu.

Présentation de l'invention

- [0023] La présente invention a pour objectif de palier les inconvénients précités.
- [0024] À cet effet, l'invention concerne un procédé de détermination du niveau d'atypicité d'individus, permettant notamment de détecter statistiquement des individus atypiques dans un jeu de données préalablement récoltées, issues de mesures de paramètres d'individus réalisées par une pluralité de systèmes de mesures.
- [0025] Le procédé comporte les étapes de :
- prétraitement des données ;
 - détermination d'un indice d'atypicité multivarié, ledit indice étant transformé par une fonction f pour être compris entre 0 et 1, sur l'ensemble des mesures pour chaque individu, à partir des données prétraitées ;
 - Identification des individus atypiques.
- [0026] Dans des modes particuliers de réalisation, l'invention répond en outre aux caractéristiques suivantes, mises en œuvre séparément ou en chacune de leurs combinaisons techniquement opérantes.
- [0027] Dans des modes de mise en œuvre de l'invention, l'étape de prétraitement est réalisée par une étape de standardisation dans laquelle pour chaque paramètre j mesuré, on centre chaque individu $x(i,j)$ par la moyenne empirique μ_j de l'ensemble des valeurs du paramètre j et on divise par l'écart-type empirique σ_j .
- [0028] Dans des modes de mise en œuvre de l'invention, l'étape de prétraitement est réalisée par une étape de standardisation robuste mettant en œuvre des indicateurs statistiques robustes.
- [0029] Dans des modes de mise en œuvre de l'invention, à la suite de l'étape de détermination, une étape d'identification et de sélection d'un sous-ensemble d'individus dont l'indice d'atypicité est nul est réalisée.
- [0030] Dans des modes de mise en œuvre de l'invention, lors de l'étape de détermination d'un indice d'atypicité multivarié, on somme les écarts en valeur absolue à la moyenne sur chaque variable p , pour chaque individu i et pour tout écart supérieur à une valeur de référence k .
- [0031] Dans des modes de mise en œuvre de l'invention, on applique aux indices bruts d'atypicité multivarié la transformation suivante, avec Φ la fonction de répartition d'une loi inverse gaussienne:
- [0032] [Math.1]
- $$z_i = \Phi \left(\frac{\text{indice}_i - \mu_{\text{indice}_i}}{\sigma_{\text{indice}_i}} \right)$$
- [0033] Dans des modes de mise en œuvre de l'invention, l'étape de détermination d'un indice d'atypicité multivarié comporte une sous-étape dans laquelle on détermine l'indice d'atypicité d'individus formant un sous-ensemble dit « sous-ensemble de

référence », et une sous-étape dans laquelle on détermine un indice d'atypicité multivarié d'un nouvel individu, l'étape d'identification des individus atypiques étant ensuite appliquée au nouvel individu au regard des indices d'atypicité des individus du sous-ensemble de référence, ledit nouvel individu étant alors intégré au sous-ensemble de référence s'il est considéré comme non atypique, et étant alors directement rejeté s'il est considéré comme atypique.

- [0034] Dans des modes de mise en œuvre de l'invention, l'étape d'identification d'individus atypiques comporte des sous étapes dans lesquelles :
- les indices d'atypicité z_i déterminés des individus sont ordonnés par ordre croissant ;
 - l'écart entre deux indices d'atypicité z_i consécutifs est mesuré ;
 - des groupes d'individus ayant des indices d'atypicité statistiquement similaires sont déterminés et
 - les individus, et plus particulièrement les groupes d'individus, atypiques sont identifiés sur la base d'un taux maximal d'individus atypiques acceptables prédéfini.
- [0035] Dans des modes de mise en œuvre de l'invention, lors de l'étape d'identification d'individus atypiques, on détermine si le groupe d'individus dans lequel les individus présentent des indices d'atypicité les plus élevés comporte un nombre d'individu moins élevé qu'un seuil maximal d'individus atypiques acceptables prédéterminé en fonction du taux maximal prédéfini. Si c'est le cas, on procède à la détermination d'un ensemble de groupes comportant le premier groupe et le ou les groupes suivants par ordre successif et décroissant, de telle sorte que la somme des individus des groupes de l'ensemble soit inférieure au seuil prédéterminé.
- [0036] Selon un autre aspect, la présente invention concerne également une utilisation du procédé tel que décrit précédemment, pour détecter des lots de vaccins atypiques dans un échantillon de lots de vaccins. C'est-à-dire que les individus sont des lots de vaccins.
- [0037] Selon un autre aspect, la présente invention concerne également une utilisation du procédé tel que décrit précédemment, pour détecter des composants électroniques atypiques dans un échantillon de composants électroniques. C'est-à-dire que les individus sont des composants électroniques.
- [0038] Selon un autre aspect, la présente invention concerne également une utilisation du procédé tel que décrit précédemment, pour détecter des mesures atypiques issues de données provenant de capteurs installés sur des équipements de production ou de mesure.
- [0039] Ainsi, dans le présent texte, le terme « individu » s'étend également à des données représentatives de mesures.
- [0040] Ces caractéristiques permettent avantageusement à l'invention de s'intégrer dans des

applications de maintenance prévisionnelle, l'atypicité d'une mesure pouvant être un indicateur d'un défaut latent de l'équipement de production ou de mesure duquel sont issues les données.

Brève description des figures

[0041] L'invention sera mieux comprise à la lecture de la description suivante, donnée à titre d'exemple nullement limitatif, et faite en se référant aux figures qui représentent :

[0042] [fig.1] un logigramme représentant les étapes du procédé selon l'invention ;

[0043] [fig.2] un graphe sur lequel est représenté un échantillon d'individus ordonnés de manière croissante en fonction de leur indice d'atypicité.

[0044] Dans ces figures, des références numériques identiques d'une figure à l'autre désignent des éléments identiques ou analogues. Par ailleurs, pour des raisons de clarté, les dessins ne sont pas à l'échelle, sauf mention contraire.

Description des modes de réalisation

[0045] Le procédé de détermination du niveau d'atypicité d'individus, permettant notamment de détecter statistiquement des individus atypiques selon la présente invention s'applique sur un jeu de données préalablement récoltées et caractérisées par une ou plusieurs mesures réalisées par des systèmes de mesures sur une pluralité d'individus.

[0046] Dans la présente demande un système de mesures peut être un testeur électrique, un système de mesure biologique, un capteur de mesure physique ou plus généralement tout outil adapté à réaliser une mesure.

[0047] Par exemple, ce jeu de données peut être organisé sous forme d'un tableau de données dont les lignes représentent n individus et dont les colonnes représentent p variables correspondant aux mesures effectuées sur les individus. Les valeurs de ces mesures sont numériques ou binaires.

[0048] Avantageusement, le jeu de données peut contenir un nombre de mesures différent pour chaque individu sans que cela soit préjudiciable au bon fonctionnement du procédé selon l'invention. En outre, le nombre de mesures peut être supérieur ou égal au nombre d'individus, et inversement, le nombre d'individus peut être supérieur ou égal au nombre de mesures, sans que cela impacte le fonctionnement du procédé selon l'invention.

[0049] Le jeu de données est issu d'un échantillon d'au moins trois individus, le nombre de variables pouvant être compris entre un et plusieurs milliers, la seule limite étant la puissance de calcul des moyens informatiques avec lesquels l'invention est destinée à être mise en œuvre.

[0050] A titre d'exemple d'application non limitatif, les individus peuvent être des composants électroniques, des lots de vaccins, etc.

- [0051] En outre, les mesures peuvent être réalisées sur des paramètres ou des caractéristiques physiques d'individus produits par des moyens de production industrielle, par exemple en fin de production, dans le cadre d'un contrôle qualité desdits individus.
- [0052] Comme le montre le logigramme de la figure 1, le procédé comporte les étapes successives de :
- prétraitement 100 des données ;
 - détermination 200 d'un indice d'atypicité multivarié, ledit indice étant transformé par une fonction f pour être compris entre 0 et 1, sur l'ensemble des mesures pour chaque individu ;
 - identification 300 des individus atypiques.
- [0053] Le prétraitement des données permet d'obtenir des données indépendantes de leur unité ou de leur échelle de mesure.
- [0054] Dans la suite du texte, on définit x_1, \dots, x_n , n observations caractérisées par p variables quantitatives. Plus précisément, dans l'exemple d'application du procédé selon l'invention, x_i représente l'ensemble des mesures effectuées sur la pièce i .
- [0055] Dans cette étape de prétraitement 100, on cherche à standardiser des données initiales $x_{i,j}$ afin d'obtenir de nouvelles données $y_{i,j}$.
- [0056] Dans un exemple de réalisation de l'invention, l'étape de prétraitement 100 peut être réalisée par une étape de standardisation non robuste pouvant être caractérisée comme décrit ci-dessous.
- [0057] Pour chaque paramètre j mesuré, on centre chaque individu $x_{i,j}$ par la moyenne empirique μ_j de l'ensemble des valeurs du paramètre j et on divise par l'écart-type empirique σ_j :
- [Math.2]
- $$y_{i,j} = \frac{x_{i,j} - \mu_j}{\sigma_j}$$
- [0058] Dans un autre exemple de réalisation de l'invention, l'étape de prétraitement 100 peut être réalisée par une étape de standardisation robuste.
- [0059] Plus précisément, l'étape de standardisation robuste peut mettre en œuvre des indicateurs statistiques robustes tels que la médiane ou la moyenne tronquée en guise de paramètre de position, c'est-à-dire à la place de la moyenne empirique μ_j dans l'équation formulée ci-avant, et de l'écart interquartile ou la déviation médiane absolue pour le paramètre d'échelle, c'est-à-dire à la place de l'écart-type empirique σ_j dans l'équation formulée ci-avant.
- [0060] Concrètement, l'étape de standardisation non robuste est préférée à l'étape de standardisation robuste en présence de données ne suivant pas une loi Normale ou lorsque les valeurs de ces données sont très extrêmes. En effet, les estimateurs non robustes comme la moyenne sont sensibles à la présence de valeurs extrêmes, contrairement à la

médiane.

[0061] L'étape de détermination 200 d'un indice d'atypicité multivarié est ensuite mise en œuvre sur l'ensemble des mesures pour chaque individu.

[0062] Plus particulièrement, dans cette étape on cherche à déterminer un indice brut d'atypicité multivarié pour chaque pièce ou individu i à partir des données prétraitées $y_{i,j}$.

[0063] A cet effet, pour chaque individu i , on somme les écarts en valeur absolue à la moyenne sur chaque variable p , pour tout écart supérieur à une valeur de référence k , avec $k \in \mathbb{R}^+$.

[Math.3]

$$indice_i = \sum_{j=1}^p \begin{cases} |y_{i,j}| - k, & \text{si } |y_{i,j}| > k \\ 0, & \text{sinon} \end{cases}$$

[0064] Les indices brut d'atypicité à valeurs dans \mathbb{R}^+ subissent une transformation f qui les projettent entre $[0,1]$. On obtient alors des indices d'atypicités z_i dont la valeur est comprise entre 0 et 1, ce qui les rend comparables les uns avec les autres.

[Math.4]

$$z_i = f(indice_i)$$

[0065] Alternativement à la transformation précédente, on peut appliquer aux indices bruts d'atypicité la transformation suivante :

[Math.5]

$$z_i = \Phi\left(\frac{indice_i - \mu_{indice_i}}{\sigma_{indice_i}}\right)$$

[0066] Avec Φ la fonction de répartition d'une loi inverse gaussienne de moyenne et de dispersion égales à 1.

[0067] Avantagusement, l'étape de détermination 200 d'un indice d'atypicité multivarié permet de créer un indice qui agrège les atypicités univariées, même faibles.

[0068] Plus particulièrement, l'indice d'atypicité permet d'identifier comme potentiellement atypiques des individus, d'une part, faiblement atypiques de manière univariée sur un grand nombre de paramètres, et d'autre part, fortement atypique sur un ou peu de paramètres.

[0069] Il y a lieu de noter que les individus dont l'indice d'atypicité est nul ne présente pas d'atypicité, et que plus un individu aura un indice d'atypicité proche de 1, plus il sera susceptible d'être atypique.

[0070] On comprend ici qu'un individu dont l'indice d'atypicité est non nul peut être soit non atypique soit atypique.

[0071] La présente invention permet donc avantagusement de détecter les individus ne présentant aucune atypicité, dans la mesure où l'indice d'atypicité d'un individu peut

être nul. Si l'ensemble des individus présente un indice d'atypicité nul, il n'y a pas d'individu atypique dans l'échantillon d'individus analysé.

- [0072] Ainsi, il est possible d'identifier et de sélectionner, à la suite de l'étape de détermination 200, un sous-ensemble d'individus dont on est sûr qu'ils ne sont pas atypiques, en identifiant et sélectionnant les individus dont l'indice d'atypicité est nul.
- [0073] Ce sous-ensemble est appelé ici « sous-ensemble de référence ». Cette sélection des « meilleurs » individus, meilleurs au sens où ils n'ont aucune atypicité, n'est pas juste le symétrique de la procédure qui consiste à éliminer tous les atypiques car cette dernière juge statistiquement du niveau d'atypicité alors que la sélection d'un sous-ensemble de référence va plus loin en éliminant des individus qui ont un peu d'atypicité même si elle n'est pas avérée statistiquement.
- [0074] La sélection d'un sous-ensemble de référence est particulièrement avantageuse, notamment car elle permet d'isoler des pièces ne présentant aucune atypicité et d'employer ces pièces. Ceci est capital dans certains types d'industrie, par exemple, dans l'industrie spatiale ou dans les grands programmes spatiaux, seules les pièces ne présentant aucune atypicité sont envoyées dans l'espace.
- [0075] Pour résumer, le procédé selon l'invention permet d'identifier des individus atypiques et des individus ne présentant pas d'atypicité.
- [0076] En outre, l'indice d'atypicité multivarié est calculable même s'il y a plus de variables que d'observations, contrairement à la plupart des méthodes de détection multivariée.
- [0077] Cette particularité est particulièrement avantageuse dans la mesure où un nombre trop important de variables au regard du nombre d'observations peut générer un bruit susceptible d'empêcher ou de complexifier l'identification d'individus atypiques.
- [0078] Enfin, ces caractéristiques permettent avantageusement de pouvoir appliquer le procédé à des jeux de données comportant des données manquantes.
- [0079] La normalisation entre 0 et 1 des indices d'atypicité, à l'issue de l'étape de détermination 200 desdits indices, confère un avantage industriel important notamment quand la traçabilité des pièces n'est pas assurée lors de leur contrôle qualité après production.
- [0080] En effet, grâce au procédé selon l'invention, dans un mode particulier de réalisation de l'invention, l'étape de détermination 200 d'un indice d'atypicité multivarié comporte une sous-étape dans laquelle on détermine un indice d'atypicité d'individus formant un sous-ensemble dit « sous-ensemble de référence », et une sous-étape dans laquelle on détermine un indice d'atypicité d'un nouvel individu.
- [0081] L'étape d'identification 300 des individus atypiques est ensuite appliquée au nouvel individu au regard des indices d'atypicité des individus du sous-ensemble de référence.
- [0082] Ledit nouvel individu est alors directement rejeté s'il est considéré comme atypique.
- [0083] De façon alternative, l'étape de détermination 200 d'un indice d'atypicité multivarié

peut comporter des sous-étape dans lesquelles on détermine l'indice d'atypicité d'individus formant le sous-ensemble de référence, puis on détermine des limites d'indices d'atypicité dudit sous-ensemble de référence sur la base des indices d'atypicité de ses individus. Par exemple, ces limites sont définies selon l'écart-type du sous-ensemble de référence, de type $\mu \pm k\sigma$ avec $k=3$ par exemple, μ et σ étant déterminés en fonction des valeurs du sous-ensemble de référence.

[0084] Ensuite, de manière analogue à ce qui a précédemment été décrit, on réalise une sous-étape de détermination d'un indice d'atypicité d'un nouvel individu.

[0085] Il est ainsi possible de réaliser un roulement du type « premier entré, premier sorti », connu de l'homme du métier sous l'acronyme anglophone « FIFO » pour « first in first out ».

[0086] Par conséquent, il est possible d'éliminer toute pièce atypique de façon intégrée au process de production industriel, en continue.

[0087] Ces modes de mise en œuvre dynamique de l'invention sont particulièrement avantageux dans les cas dans lesquels lors de la production la traçabilité des individus produits n'est pas assurée physiquement, et que les individus atypiques détectées avec les méthodes conventionnelles de détection d'individus atypiques, dites « post-processing », ne peuvent pas être rebutés.

[0088] Les méthodes de post-processing consistent à identifier statistiquement les individus atypiques à partir d'un échantillon d'individus, ensuite d'identifier physiquement les individus atypiques pour les éliminer. Ceci nécessite donc la traçabilité de chaque individu produit et de traiter un échantillon d'individus. La présente invention lève cette contrainte et permet d'éliminer les individus à la volée.

[0089] Il y a lieu de noter que grâce à la normalisation des indices d'atypicité, celui du nouvel individu observé est comparable à celui des autres individus, même s'il est issu d'un nouveau lot ce qui n'est pas forcément le cas avec les méthodes de détection statistiques connues.

[0090] L'étape d'identification 300 des individus atypiques peut consister préférentiellement à mettre en œuvre les sous-étapes suivantes :

- ordonner les indices d'atypicité z_i déterminés des individus par ordre croissant, comme le montre le graphe de la figure 2, soit :

[Math.6]

$$z_{(1)} \leq \dots \leq z_{(n)}$$

- mesurer l'écart entre deux indices d'atypicité z_i consécutifs :

[Math.7]

$$w_i = z_{(i)} - z_{(i-1)}$$

Ces différences w_i suivent la loi de distribution suivante :

[Math.8]

$$F(w) = 1 - (1 - w)^n$$

de moyenne

[Math.9]

$$\frac{1}{n+1}$$

- identifier les individus atypiques par rapport à l'ensemble des individus.

[0091] Cette dernière sous-étape est réalisée par identification de groupes d'individus ayant des indices d'atypicité statistiquement similaires. Cette identification est réalisée selon les écarts entre deux indices d'atypicité successifs.

[0092] Dans le graphe de la figure 2, les groupes d'individus sont séparés par des traits discontinus horizontaux.

[0093] Plus particulièrement, les écarts entre deux indices d'atypicité successifs supérieurs au α -ième percentile sont identifiés :

[Math.10]

$$v(\alpha) = 1 - (1 - \alpha)^{\frac{1}{n}}$$

[0094] Dans l'exemple de réalisation représenté sur la figure 2, la valeur de percentile choisie est $\alpha = 90\%$.

[0095] On calcule ces écarts, car théoriquement les différences des écarts suivent la loi de distribution décrite précédemment en [Math.7], et le procédé peut s'appuyer sur un seuil théorique pour identifier des groupes d'individus.

[0096] Ainsi, une comparaison entre individus est avantageusement réalisée de sorte que si un individu est statistiquement atypique mais pas significativement beaucoup plus que les autres individus alors il n'est pas identifié comme atypique.

[0097] Grâce à cette opération d'identification, des groupes d'individus au niveau d'atypicité sensiblement semblables ou analogues sont déterminés. De cette manière l'ensemble des individus formant un groupe est identifié comme atypique ou comme non atypique. Ainsi, la fiabilité de la détection des individus atypique est considérablement augmentée dans la mesure où le procédé garanti que deux individus ayant un indice d'atypicité très proche sont tous les deux identifiés comme atypiques ou comme non atypiques. Et il est donc possible d'aller au-delà de la détection d'individus atypiques, puisqu'il est ainsi possible de générer un sous-ensemble d'individus présentant des indices d'atypicité homogènes.

[0098] Avantageusement, l'approche du procédé de la présente invention est ainsi d'identifier des groupes d'individus atypiques.

[0099] Les groupes d'individus au niveau d'atypicité sensiblement semblables ou analogues étant déterminés et les indices d'atypicité z_i déterminés des individus étant ordonnés

par ordre croissant, les groupes sont, de fait, ordonnés par ordre croissant suivant le niveau d'atypicité de leurs individus respectifs, comme le montre la figure 2.

- [0100] L'étape suivante du procédé est l'identification des groupes d'individus atypiques en fonction d'un taux maximal d'individus atypiques acceptables prédéterminé.
- [0101] Dans l'exemple de réalisation représenté sur la figure 2, le taux maximal d'individus atypiques acceptables est de 10%. Le graphe représentant 57 individus, le seuil maximal d'individus atypiques acceptables est de 5.
- [0102] Plus spécifiquement, lors de cette étape, on détermine si le groupe d'individus dans lequel les individus présentent des indices d'atypicité les plus élevés, dit « premier groupe », comporte un nombre d'individu inférieur ou égal au seuil maximal d'individus atypiques acceptables prédéterminé.
- [0103] Si c'est le cas, alors tous les individus de ce groupe sont identifiés comme atypiques, sinon ils sont considérés comme non atypiques.
- [0104] Si les individus du premier groupe sont identifiés comme atypiques, on procède à l'identification du groupe dont le ou les individus présentent des indices d'atypicité immédiatement inférieurs et chacun des groupes suivants, de façon successive, tant que la somme des individus dudit groupe ou de l'ensemble desdits groupes est inférieure ou égale au seuil maximal, les individus dudit ou desdits groupes identifiés étant alors identifiés comme atypiques.
- [0105] En d'autres termes, si les individus du premier groupe sont identifiés comme atypiques, on cherche à savoir si ceux du groupe dans lequel les individus présentent des indices d'atypicité immédiatement inférieur à ceux du premier groupe, dit « deuxième groupe » le sont également.
- [0106] Dans l'exemple représenté sur la figure 2, le premier groupe comporte un seul individu, la recherche d'individus atypiques est donc étendue au groupe suivant, c'est-à-dire au deuxième groupe.
- [0107] On additionne alors ensuite le nombre d'individus des premier et deuxième groupes.
- [0108] Si la somme des individus des premier et deuxième groupes est inférieure ou égale au seuil maximal, alors tous les individus du deuxième groupe sont également identifiés comme atypiques, sinon ils sont considérés comme non atypiques.
- [0109] Dans l'exemple représenté sur la figure 2, le deuxième groupe comportant trois individus, la somme d'individus des premier et deuxième groupes est égale à quatre, la recherche d'individus atypique est étendue au groupe suivant, c'est-à-dire au troisième groupe. Les individus des premiers et deuxième groupe sont considérés comme atypiques.
- [0110] On itère ces opérations jusqu'à identifier un groupe dont les individus sont non atypiques, c'est-à-dire, jusqu'à ce que la somme des individus des groupes étudiés soit supérieure au seuil maximal. A partir de ce groupe, tous les groupes suivants,

c'est-à-dire les groupes présentant des individus dont les indices d'atypicité sont inférieurs à ceux des groupes étudiés, sont donc constitués d'individus non atypiques.

- [0111] Dans l'exemple représenté sur la figure 2, comme le troisième groupe comporte un unique individu, la somme d'individus du premier au troisième groupes est égale à cinq. Les individus du premier au troisième groupes sont considérés comme atypiques, les individus des autres groupes étant considérés comme non atypiques.
- [0112] Avantagusement, grâce à ces opérations, le taux de fausses alertes, c'est-à-dire d'erreur dans l'identification d'individus atypiques, est sensiblement réduit, et ce, sans nécessiter d'imposer un taux strict d'individus atypiques à respecter.
- [0113] Une fois détecté, les individus atypiques peuvent enfin être identifiés grâce à une traçabilité physique, ou bien être éliminés directement grâce au mode de mise en œuvre dynamique.
- [0114] Alternativement, dans un autre mode de mise en œuvre du procédé selon l'invention, l'étape d'identification 300 des individus atypiques peut consister en l'application d'une méthode statistique sur l'indice d'atypicité pour détecter les individus atypiques.
- [0115] Par exemple, on peut calculer une limite de type $\mu \pm k\sigma$ avec $k=3$ par exemple, μ et σ étant déterminés en fonction des valeurs de l'échantillon. Ces individus atypiques sont alors éliminés de la production, si le processus de production des individus permet la traçabilité des individus produits.
- [0116] La présente invention permet alors de revenir à un cas univarié dans la mesure où il n'y a qu'un seul indice d'atypicité par individu, mais en éliminant ou en limitant considérablement les fausses alarmes par rapport aux méthodes de détection statistiques de l'état de l'art.
- [0117] De manière plus générale, il est à noter que les modes de mise en œuvre et de réalisation de l'invention considérés ci-dessus ont été décrits à titre d'exemples non limitatifs et que d'autres variantes sont par conséquent envisageables.

Revendications

- [Revendication 1] Procédé de détermination du niveau d'atypicité d'individus, permettant notamment de détecter statistiquement des individus atypiques dans un jeu de données préalablement récoltées, issues de mesures de paramètres d'individus réalisées par une pluralité de systèmes de mesures, caractérisé en ce qu'il comporte les étapes de :
- prétraitement 100 des données ;
 - détermination 200 d'un indice d'atypicité multivarié, ledit indice étant transformé par une fonction f pour être compris entre 0 et 1, sur l'ensemble des mesures pour chaque individu, à partir des données pré-traitées ;
 - Identification 300 des individus atypiques.
- [Revendication 2] Procédé selon la revendication 1, dans lequel l'étape de prétraitement est réalisée par une étape de standardisation dans laquelle pour chaque paramètre j mesuré, on centre chaque individu $x(i,j)$ par la moyenne empirique μ_j de l'ensemble des valeurs du paramètre j et on divise par l'écart-type empirique σ_j .
- [Revendication 3] Procédé selon la revendication 1, dans lequel l'étape de prétraitement 100 est réalisée par une étape de standardisation robuste mettant en œuvre des indicateurs statistiques robustes.
- [Revendication 4] Procédé selon l'une des revendications 1 à 3, dans lequel à la suite de l'étape de détermination 200, une étape d'identification et de sélection d'un sous-ensemble d'individus dont l'indice d'atypicité est nul est réalisée.
- [Revendication 5] Procédé selon l'une des revendications 1 à 4, dans lequel lors de l'étape de détermination 200 d'un indice d'atypicité multivarié, on somme les écarts en valeur absolue à la moyenne sur chaque variable p , pour chaque individu i et pour tout écart supérieur à une valeur de référence k afin de déterminer un indice brut d'atypicité multivarié.
- [Revendication 6] Procédé selon la revendication 5, dans lequel on applique aux indices bruts d'atypicité multivarié la transformation suivante, avec Φ la fonction de répartition d'une loi inverse gaussienne:
- [Math.11]
- $$z_i = \Phi \left(\frac{\text{indice}_i - \mu_{\text{indice}_i}}{\sigma_{\text{indice}_i}} \right)$$
- [Revendication 7] Procédé selon l'une des revendications 1 à 6, dans lequel l'étape de détermination 200 d'un indice d'atypicité multivarié comporte une sous-

étape dans laquelle on détermine un indice d'atypicité d'individus formant un sous-ensemble dit « sous-ensemble de référence », et une sous-étape dans laquelle on détermine un indice d'atypicité multivarié d'un nouvel individu, l'étape d'identification 300 des individus atypiques étant ensuite appliquée au nouvel individu au regard des indices d'atypicité des individus du sous-ensemble de référence, ledit nouvel individu étant alors intégré au sous-ensemble de référence s'il est considéré comme non atypique, et étant rejeté s'il est considéré comme atypique.

[Revendication 8]

Procédé selon l'une des revendications 1 à 7, dans lequel l'étape d'identification 300 d'individus atypiques comporte des sous étapes dans lesquelles :

- les indices d'atypicité z_i déterminés des individus sont ordonnés par ordre croissant ;
- l'écart entre deux indices d'atypicité z_i consécutifs est mesuré ;
- des groupes d'individus ayant des indices d'atypicité statistiquement similaires sont déterminés et
- les individus atypiques sont identifiés sur la base d'un taux maximal d'individus atypiques acceptables prédéfini.

[Revendication 9]

Procédé selon la revendication 8, dans lequel lors de l'étape d'identification 300 d'individus atypiques, on détermine si le groupe d'individus dans lequel les individus présentent des indices d'atypicité les plus élevés comporte un nombre d'individu moins élevé qu'un seuil maximal d'individus atypiques acceptables prédéterminé en fonction du taux maximal prédéfini, si c'est le cas, on procède à la détermination d'un ensemble de groupes comportant le premier groupe et le ou les groupes suivants par ordre successif et décroissant, de telle sorte que la somme des individus des groupes de l'ensemble soit inférieure au seuil prédéterminé.

[Revendication 10]

Utilisation du procédé selon l'une des revendications 1 à 9, pour détecter des lots de vaccins atypiques dans un échantillon de lots de vaccins.

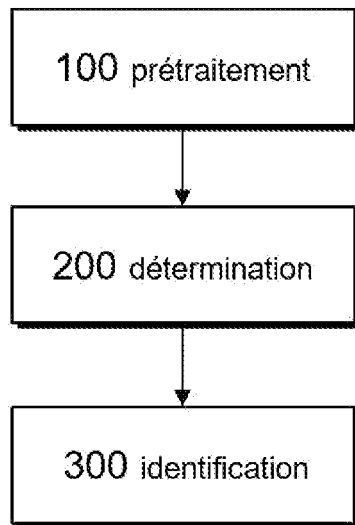
[Revendication 11]

Utilisation du procédé selon l'une des revendications 1 à 9, pour détecter des composants électroniques atypiques dans un échantillon de composants électroniques.

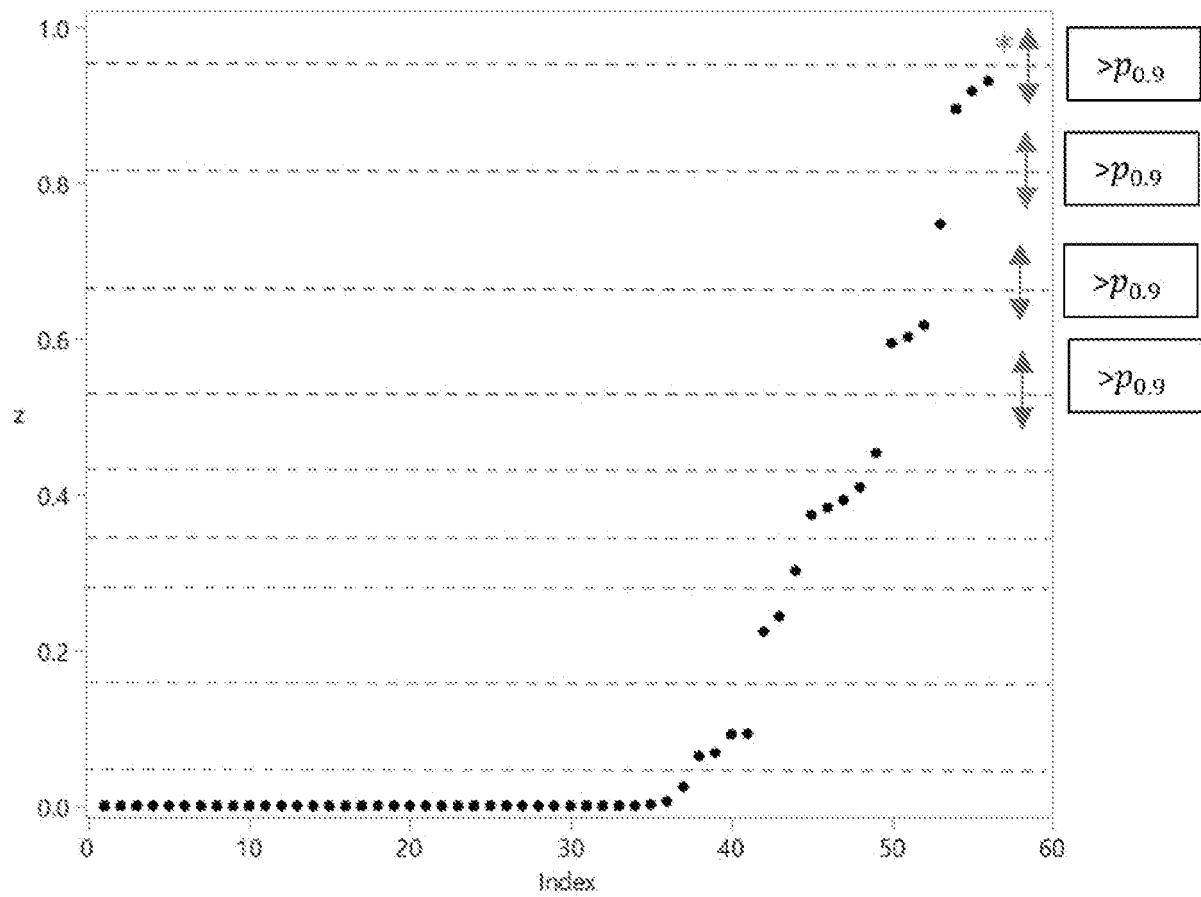
[Revendication 12]

Utilisation du procédé selon l'une des revendications 1 à 9, pour détecter des mesures atypiques issues de données provenant de capteurs installés sur des équipements de production ou de mesure.

[Fig. 1]



[Fig. 2]



**RAPPORT DE RECHERCHE
PRÉLIMINAIRE**

établi sur la base des dernières revendications
déposées avant le commencement de la recherche

N° d'enregistrement
national

FA 885898
FR 2008243

DOCUMENTS CONSIDÉRÉS COMME PERTINENTS		Revendication(s) concernée(s)	Classement attribué à l'invention par l'INPI
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes		
X	CHRISTOPHER J TRALIE ET AL: "Multi-scale Geometric Summaries for Similarity-based Sensor Fusion", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 13 octobre 2018 (2018-10-13), XP081069712, * sections 1, 4, 5 * -----	1-12	G06F17/10 G06F16/2458 A61K39/00
X	SHUCHU HAN ET AL: "Supervised Feature Subset Selection and Feature Ranking for Multivariate Time Series without Feature Extraction", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 1 mai 2020 (2020-05-01), XP081656918, * sections 1, 3 * -----	1-12	
			DOMAINES TECHNIQUES RECHERCHÉS (IPC)
			G06F
Date d'achèvement de la recherche		Examineur	
19 mars 2021		Virnik, Elena	
CATÉGORIE DES DOCUMENTS CITÉS		T : théorie ou principe à la base de l'invention	
X : particulièrement pertinent à lui seul		E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure.	
Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie		D : cité dans la demande	
A : arrière-plan technologique		L : cité pour d'autres raisons	
O : divulgation non-écrite		
P : document intercalaire		& : membre de la même famille, document correspondant	