

⑫

CERTIFICAT D'UTILITÉ

B3

⑤④ Détection d'anomalies dans les données billettiques de transport en commun.

②② Date de dépôt : 15.12.22.

③③ Priorité :

⑥⑥ Références à d'autres documents nationaux
apparentés :

☐ Demande(s) d'extension :

⑦① Demandeur(s) : *Revenue Collection Systems France
SAS Société par actions simplifiée à associé unique —
FR.*

④③ Date de mise à la disposition du public
de la demande : 21.06.24 Bulletin 24/25.

④⑤ Date de la mise à disposition du public du
certificat d'utilité : 17.01.25 Bulletin 25/03.

⑤⑥ Les certificats d'utilité ne font pas l'objet d'un
rapport de recherche.

⑦② Inventeur(s) : DUGUET Stéphane, LADJELATE
Bilal et LAMPE Rodolphe.

⑦③ Titulaire(s) : Revenue Collection Systems France
SAS Société par actions simplifiée à associé unique.

⑦④ Mandataire(s) : Lavoix.

FR 3 143 802 - B3



Description

Titre de l'invention : Détection d'anomalies dans les données billettiques de transport en commun

- [0001] La présente invention concerne un procédé et un système de détection d'anomalies dans les données billettiques de transport en commun.
- [0002] L'invention concerne également un programme d'ordinateur comportant des instructions logicielles qui, lorsqu'elles sont exécutées par un ordinateur, mettent en œuvre un tel procédé de détection d'anomalies dans les données billettiques de transport en commun.
- [0003] Le domaine de l'invention porte sur la détection d'anomalie pour des systèmes billettiques permettant aux usagers munis d'un titre de transport d'accéder aux réseaux de transports publics.
- [0004] Plus précisément, de tels systèmes billettiques comprennent notamment les guichets et machines de vente, les valideurs de titres et les portillons automatiques de contrôle d'accès, et créent de grandes quantités de données à la fois quantitatives et qualitatives sur les actions des passagers sur le ou les réseau(x) de transports publics, notamment relatives aux validations, aux achats de produit, au rechargement, et aux remboursements, et de différents types portant par exemple sur des valeurs de transactions, des balises entrées ou sorties, des stations d'arrêts.
- [0005] A partir de ces données, des traitements automatiques permettent de rechercher les anomalies et fraudes potentielles qui peuvent avoir lieu au cours des transactions des usagers du réseau de transport.
- [0006] De tels traitements automatiques existants dédiés au contrôle de données billettiques de transport en commun visent généralement à appliquer des règles prédéfinies aux transactions pour trouver des phénomènes impossibles ou douteux ou correspondent
- [0007] à des traitements de données statistiques, et de façon extensive pour détecter des points hors distribution. Ces solutions actuelles présentent l'inconvénient d'être peu adaptées à la découverte de nouvelles anomalies ou de nouvelles fraudes, car nécessitant un traitement souvent très lourd, complexe et peu évolutif car nécessitant une réflexion humaine avancée, par un opérateur de supervision, sur ce qui constitue une fraude ou une anomalie.
- [0008] Dans d'autres domaines d'applications différentes des méthodes de traitement par apprentissage qui présentent l'intérêt d'être légères à déployer, et de découvrir de nouveaux schémas de fraude qui ne sont pas forcément accessibles aux méthodes de traitement statistiques sont utilisées mais non adaptées au contrôle de données billettiques de transport en commun. Les nouveaux schémas de fraude nécessitent un

apprentissage non-supervisé ou auto-supervisé, car il n'est pas possible de savoir a priori si un comportement est anormal, sinon il aurait été possible de l'intégrer dans un algorithme de traitement statistique.

- [0009] A l'heure actuelle, il n'existe donc aucun traitement automatique propre à découvrir de nouveaux schémas de fraude dans le domaine du contrôle de données billettiques de transport en commun.
- [0010] A cet effet, l'invention a pour objet un procédé de détection d'anomalies dans les données billettiques de transport en commun, le procédé étant propre à être mis en oeuvre par un système électronique de détection d'anomalies dans les données billettiques de transport en commun, le procédé comprenant au moins une phase d'apprentissage et au moins une phase d'inférence,
- [0011] ladite au moins une phase d'apprentissage comprenant les étapes suivantes, appliquées à des données billettiques tabulaires d'entraînement :
- [0012] - prétraitement et sélection de colonnes de données d'intérêt prédéterminées au sein desdites données billettiques tabulaires d'entraînement ;
- [0013] - apprentissage profond d'au moins un réseau de neurones en utilisant les données billettiques d'entraînement optimisées par ledit prétraitement et obtention d'un modèle dudit au moins un réseau de neurones configuré pour détecter une anomalie de données billettiques ;
- [0014] ladite au moins une phase d'inférence comprenant les étapes suivantes, appliquées à des données billettiques tabulaires de test ;
- [0015] - application dudit prétraitement et de ladite sélection de colonnes de données d'intérêt prédéterminées auxdites données billettiques tabulaires de test ;
- [0016] - obtention d'un score correspondant à un niveau d'anomalie desdites données billettiques tabulaires de test par application dudit modèle auxdites données billettiques de test optimisées par ledit prétraitement.
- [0017] Ainsi, le procédé selon la présente invention permet d'accélérer les traitements de données billettiques de transport en commun tout en recherchant des comportements anormaux rares et complexes.
- [0018] Plus précisément, le procédé selon la présente invention propose un prétraitement spécifique au format de données tabulaires utilisé dans les système billettiques de transport en commun afin d'optimiser les entrées d'un réseau de neurones dont le modèle, après avoir été entraîné à la détection d'anomalies à partir de données d'entraînement en utilisant un apprentissage profond avec une perte contrastive, est apte à être utilisé pour fournir un score propre à décrire quantitativement à quel point un évènement billettique est anormal ou non.
- [0019] Suivant d'autres aspects avantageux de l'invention, le procédé de détection d'anomalies dans les données billettiques de transport en commun comprend une ou

plusieurs des caractéristiques suivantes, prises isolément ou suivant toutes les combinaisons techniquement possibles :

- [0020] - le prétraitement et la sélection de colonnes de données d'intérêt prédéterminées comprennent :
- [0021] - la sélection, pour chaque utilisateur de transport en commun, ou par extension pour chaque unique identifiant d'un titre de transport en commun ou d'un moyen de paiement grâce auquel cette personne voyage, d'un nombre prédéterminé de dernières transactions billettiques et des données billettiques tabulaires associées ;
- [0022] - au sein desdites données billettiques tabulaires associées au nombre prédéterminé de dernières transactions, la sélection des colonnes de données correspondant au :
 - [0023] - type de transaction ;
 - [0024] - la valeur numérique de transaction ;
 - [0025] - le contexte billettique de transaction ;
- [0026] - le prétraitement et la sélection de colonnes de données d'intérêt prédéterminées comprennent en outre un formatage desdites données billettiques sélectionnées selon un format prédéterminé ;
- [0027] - le prétraitement et la sélection de colonnes de données d'intérêt prédéterminées comprennent en outre un tri des transactions dans l'ordre chronologique suivi d'un groupement desdites transactions par utilisateur, chaque utilisateur étant identifié par un identifiant unique, puis un filtrage des utilisateurs afin de ne conserver que les données billettiques associées aux utilisateurs ayant un nombre de transactions supérieur ou égal audit nombre prédéterminé de dernières transactions billettiques ;
- [0028] - le prétraitement et la sélection de colonnes de données d'intérêt prédéterminées comprennent en outre un typage puis une séparation des données billettiques tabulaires sélectionnées en deux types de variables comprenant des variables catégorielles propres à décrire qualitativement un état et/ou un élément billettique et des variables numériques propres à décrire quantitativement un état et/ou un élément billettique ;
- [0029] - ledit apprentissage profond utilise une fonction de perte contrastive et dans lequel l'obtention dudit score en phase d'inférence comprend une normalisation en utilisant les valeurs minimales et maximales de la fonction de perte contrastive rencontrées en phase d'apprentissage ;
- [0030] - ledit apprentissage profond comprend :
- [0031] - la détermination, au moyen dudit au moins un réseau de neurones, d'une pluralité de vues des données billettiques issues dudit prétraitement et de ladite sélection,
- [0032] - la réduction dimensionnelle de chacune des vues de ladite pluralité de vues fournissant des vues dans un espace latent de dimension réduite ;
- [0033] - le calcul de propriétés géométriques prédéterminées dans ledit espace latent de plus petite dimension obtenue par ladite réduction dimensionnelle ;

- [0034] - ladite au moins une phase d'inférence comprend en outre une étape d'explication attribuant des poids aux colonnes sélectionnées lors de l'étape de prétraitement, lesdits poids étant propres à expliquer au moins en partie ledit score obtenu.
- [0035] L'invention a également pour objet un programme d'ordinateur comportant des instructions logicielles qui, lorsqu'elles sont exécutées par un ordinateur, mettent en œuvre un procédé de détection d'anomalies dans les données billettiques de transport en commun, tel que défini ci-dessus.
- [0036] L'invention concerne également un système de détection d'anomalies dans les données billettiques de transport en commun, le système électronique comprenant une unité d'apprentissage et une unité d'inférence,
- [0037] l'unité d'apprentissage comprenant, pour application à des données billettiques tabulaires d'entraînement :
- [0038] - un module de prétraitement configuré pour prétraiter et sélectionner des colonnes de données d'intérêt prédéterminées au sein desdites données billettiques tabulaires d'entraînement ;
- [0039] - un module d'apprentissage profond configuré pour mettre en œuvre un apprentissage profond d'au moins un réseau de neurones en utilisant les données billettiques d'entraînement optimisées par ledit prétraitement et pour obtenir un modèle dudit au moins un réseau de neurones propre à détecter une anomalie de données billettiques ;
- [0040] ladite unité d'inférence comprenant, pour application à des données billettiques tabulaires de test :
- [0041] - un module d'application configuré pour appliquer ledit prétraitement et ladite sélection de colonnes de données d'intérêt prédéterminées auxdites données billettiques tabulaires de test ;
- [0042] - un module d'obtention d'un score correspondant à un niveau d'anomalie desdites données billettiques tabulaires de test, ledit module d'obtention étant configuré pour obtenir ledit score par application dudit modèle auxdites données billettiques de test optimisées par ledit prétraitement.
- [0043] Ces caractéristiques et avantages de l'invention apparaîtront plus clairement à la lecture de la description qui va suivre, donnée uniquement à titre d'exemple non limitatif, et faite en référence aux dessins annexés, sur lesquels :
- [0044] - [Fig.1] la [Fig.1] est un schéma illustrant un système électronique de détection d'anomalies dans les données billettiques de transport en commun ;
- [0045] - [Fig.2] la [Fig.2] est un organigramme d'un procédé de détection d'anomalies dans les données billettiques de transport en commun selon la présente invention ;
- [0046] - [Fig.3] la [Fig.3] illustre en détail l'étape de prétraitement de données et sélection de colonnes de données d'intérêt prédéterminées au sein desdites données billettiques

tabulaires.

- [0047] La [Fig.1] est une représentation d'un système électronique 10 de détection d'anomalies dans les données billettiques de transport en commun selon la présente invention.
- [0048] Le système électronique 10 comprenant une unité d'apprentissage et une unité d'inférence, toutes deux non représentées à des fins de simplicité sur la [Fig.1], ces deux unités pouvant partager, selon un mode de réalisation, au moins un même module électronique.
- [0049] L'unité d'apprentissage comprend tout d'abord, pour application à des données 12 billettiques tabulaires, notées D_T, et en l'occurrence, pendant l'apprentissage, des données d'entraînement, dont il est connu au préalable si elles sont ou non porteuses d'anomalie(s), un module 14 de prétraitement configuré pour prétraiter et sélectionner (noté P+S) des colonnes de données d'intérêt prédéterminées au sein desdites données billettiques tabulaires d'entraînement.
- [0050] Le fonctionnement d'un tel module 14 de prétraitement et de sélection sera détaillé par la suite en relation avec la [Fig.3]. Un tel module 14 de prétraitement et de sélection fournit en sortie des données 16 optimisées (notées D_O_A) pour un apprentissage profond subséquent.
- [0051] L'unité d'apprentissage comprend par ailleurs un module 20 d'apprentissage profond configuré pour mettre en œuvre un apprentissage profond d'au moins un réseau de neurones en utilisant les données 16 billettiques d'entraînement optimisées par ledit module 14 de prétraitement et pour obtention d'un modèle 22 dudit au moins un réseau de neurones propre à détecter une anomalie de données billettiques.
- [0052] L'unité d'inférence comprend quant à elle, pour application à des données billettiques tabulaires D_T de test, tout d'abord un module 23 d'application en inférence (notée A_I) configuré pour appliquer ledit prétraitement et ladite sélection de colonnes de données d'intérêt prédéterminées 230 auxdites données billettiques tabulaires de test.
- [0053] Autrement dit le module 23 est configuré pour appliquer les mêmes prétraitement et sélection aux données billettiques tabulaires de test, qu'appliqués auparavant aux données billettiques d'entraînement.
- [0054] Selon une variante non représentée, l'unité d'inférence réutilise le module 14 de l'unité d'apprentissage et non un module distinct 23 tel que représenté sur la [Fig.1].
- [0055] L'unité d'inférence comprend en outre un module 24 d'obtention d'un score 26 correspondant à un niveau d'anomalie desdites données billettiques tabulaires de test, ledit module d'obtention 24 étant configuré pour obtenir ledit score 26 par application dudit modèle 22 de réseau(x) de neurones auxdites données billettiques de test optimisées par ledit prétraitement mis en œuvre par le module 23 précité.

- [0056] En complément optionnel, le système électronique 10 comprend en outre un module 28 d'explicabilité configuré pour attribuer des poids aux colonnes sélectionnées par le module 23 de prétraitement, lesdits poids étant propres à expliquer au moins en partie ledit score 26 obtenu.
- [0057] En fonction du score 26, les données billettiques tabulaires de test sont catégorisées par décision comme représentatives d'un évènement anormal 30 (i.e. une anomalie E_A), ou en prenant en compte également le modèle 22 selon la flèche 31, d'un évènement normal 32 (i.e. absence d'anomalie E_N).
- [0058] Selon une mise en œuvre non représentée, le système électronique 10 de détection d'anomalies dans les données billettiques de transport en commun selon la présente invention comprend une unité de traitement formée par exemple d'une mémoire et d'un processeur associé à la mémoire, et le système électronique 10 de détection d'anomalies dans les données billettiques de transport en commun selon la présente invention est au moins en partie réalisé sous forme d'un logiciel, ou d'une brique logicielle, exécutable par le processeur, notamment pour l'unité d'apprentissage : le module 14 de prétraitement et le module 20 d'apprentissage profond, et pour l'unité d'inférence le module d'application 23 et le module d'obtention 24. La mémoire du système électronique 10 est alors apte à stocker de tels logiciels ou briques logicielles, et le processeur est alors apte à les exécuter.
- [0059] Selon une autre mise en œuvre également non représentée, pour l'unité d'apprentissage : le module 14 de prétraitement et le module 20 d'apprentissage profond, et pour l'unité d'inférence le module d'application 23 et le module d'obtention 24 sont réalisés chacun sous forme d'un composant logique programmable, tel qu'un FPGA (de l'anglais *Field Programmable Gate Array*), ou encore sous forme d'un circuit intégré dédié, tel qu'un ASIC (de l'anglais *Application Specific integrated Circuit*).
- [0060] Lorsqu'une partie du système électronique 10 de détection d'anomalies dans les données billettiques de transport en commun selon la présente invention est réalisée sous forme d'un ou plusieurs logiciels, c'est-à-dire sous forme d'un programme d'ordinateur, cette partie est en outre apte à être enregistrée sur un support, non représenté, lisible par ordinateur. Le support lisible par ordinateur est par exemple, un médium apte à mémoriser des instructions électroniques et à être couplé à un bus d'un système informatique. A titre d'exemple, le support lisible est un disque optique, un disque magnéto-optique, une mémoire ROM, une mémoire RAM, tout type de mémoire non-volatile (par exemple EPROM, EEPROM, FLASH, NVRAM), une carte magnétique ou encore une carte optique. Sur le support lisible est alors mémorisé un programme d'ordinateur comportant des instructions logicielles.
- [0061] Le fonctionnement du système électronique 10 de détection d'anomalies dans les

données billettiques de transport en commun selon la présente invention va par la suite être décrit plus en détail en relation avec les figures 2 et 3.

- [0062] La [Fig.2] organigramme d'un procédé 34 de détection d'anomalies dans les données billettiques de transport en commun selon la présente invention mis en œuvre par le système électronique de la [Fig.1].
- [0063] Sur la [Fig.2], le procédé de détection d'anomalies dans les données billettiques de transport en commun comprend au moins une phase d'apprentissage P_1 et au moins une phase d'inférence P_2 .
- [0064] La phase d'apprentissage P_1 consiste à apprendre à détecter des anomalies à partir de données billettiques tabulaires d'entraînement (i.e. des données billettiques tabulaires d'entraînement D_T_A) dont on ne sait pas au préalable si elles sont représentatives d'au moins une anomalie ou non (entraînement non-supervisé).
- [0065] Ladite au moins une phase d'apprentissage P_1 comprend tout d'abord une étape 36 de prétraitement (noté P_A pour prétraitement en apprentissage) et sélection de colonnes de données d'intérêt prédéterminées au sein desdites données billettiques tabulaires d'entraînement D_T_A .
- [0066] Puis, la phase d'apprentissage P_1 comprend une étape 38 d'apprentissage profond d'au moins un réseau de neurones en utilisant les données billettiques d'entraînement optimisées par ledit prétraitement et d'obtention d'un modèle M dudit au moins un réseau de neurones configuré pour détecter une anomalie de données billettiques.
- [0067] Par exemple, le ou les réseau de neurone(s) utilisés en apprentissage profond puis en inférence par la suite sont des réseaux de neurone(s) convolutif(s).
- [0068] En complément optionnel, ledit apprentissage profond 38 utilise une fonction de perte contrastive.
- [0069] Selon un autre complément optionnel non représenté, ledit apprentissage profond 38 comprend :
- [0070] - la détermination, au moyen dudit au moins un réseau de neurones, d'une pluralité de vues des données billettiques issues dudit prétraitement et de ladite sélection 36,
- [0071] - la réduction dimensionnelle de chacune des vues de ladite pluralité de vues par un encodeur prédéterminé fournissant des vues dans un espace latent de dimension réduite ;
- [0072] - le calcul de propriétés géométriques prédéterminées dans ledit espace latent de plus petite dimension obtenue par ladite réduction dimensionnelle.
- [0073] Autrement dit, ce complément optionnel revient à adapter spécifiquement pour la détection d'anomalies dans les données billettiques de transport en commun, un apprentissage profond tel que décrit par Qiu et al. dans l'article intitulé « *Neural Transformation Learning for Deep Anomaly Detection Beyond Images* » de 2021.
- [0074] Plus précisément, lors dudit apprentissage profond 38, on considère X l'espace dans

lequel se trouve les données que l'on cherche à caractériser. K transformations $T_1, \dots, T_K$ où chaque transformation T_i correspond à un des réseaux de neurones travaillant sur X . L'encodeur mettant en œuvre la réduction dimensionnelle précitée est représenté par la fonction f_ϕ allant de X vers $\mathbb{R}^m$ dont la dimension m est telle que : $m < \dim(X)$. De plus, pour chaque entrée x , tel que $x_k = T_k(x)$ la fonction de similarité suivante est utilisée : $s(x_k, x_1) = \text{sim}(f_\phi(x_k), f_\phi(x_1))$, où sim est la similarité cosinus: $\text{sim}(z, z') := \frac{z^T z'}{\|z\| \|z'\|}$, de même que h la fonction de score de similarité permettant à la transformation de la similarité à valeurs dans $[-1; 1]$, de prendre des valeurs dans $\mathbb{R}$: $h(x_k, x_1) = \exp\left(\frac{s(x_k, x_1)}{\tau}\right)$, avec τ est un paramètre de « température ».

[0075] Ainsi, entre deux transformations x_k et x_1 , en utilisant la fonction h une mesure de leur score de similarité est obtenue en considérant notamment que si $h(x_k, x_1)$ est grand; alors les deux transformations sont proches et que si $h(x_k, x_1)$ est faible, les deux transformations encodées sont très lointaines, ce qui permet alors de créer la fonction L de perte contrastive propre à guider l'apprentissage des réseaux de neurones $T_1, \dots, T_K$ et de f_ϕ : tel que $L(x) = -\sum_k \log \frac{h(x_k, x)}{h(x_k, x) + \sum_{i \neq k} h(x_k, x_i)}$.

[0076] Minimiser la fonction L de perte contrastive, vise à rendre $h(x_k, x)$ aussi grand que possible, car chaque transformation associée respectivement à un réseau de neurones doit être aussi proche que possible des données d'origine selon une notion classique de proximité sémantique et ce tout en considérant que chaque score de similarité $h(x_k, x_1)$ doit être négativement aussi grand que possible, c'est-à-dire que les transformations doivent être aussi différentes que possibles selon une notion classique de diversité.

[0077] Par exemple, selon la présente invention, la structure des réseaux de neurones utilisée en terme de couche(s) de transformation est telle qu'elle combine :

[0078] - une couche convolutionnelle à une dimension sur une fenêtre de trois éléments, avec un décalage d'un élément et un biais nul,

[0079] - une couche de norme instancielle telle que $y = \frac{x - E[x]}{\sqrt{\text{Var}[x]}} * \gamma$, où seul le paramètre γ

peut changer, en considérant qu'une couche d'un réseau de neurones reçoit, lors de l'entraînement, un ensemble de points de $\mathbb{R}^n$ (n étant la dimension d'entrée), x correspondant à l'ensemble des valeurs du point de $\mathbb{R}^n$ considéré (il s'agit donc d'un vecteur de n valeurs), $E[x]$ correspondant à la moyenne et $\text{Var}[x]$ à la variance, la valeur y étant la sortie associée à x l'ensemble des valeurs du point de $\mathbb{R}^n$ considéré, et

[0080] - une fonction d'activation sigmoïde.

[0081] Selon un exemple particulier de la présente invention, le nombre de transformations

K à apprendre est fixé à seize. Une suite de couches convolutionnelle allant de trente-deux à deux-cent-cinquante-six canaux, doublant de couches en couches et s'arrêtant lorsque la série en entrée atteint une taille de huit éléments est appliquée. Une convolution à une dimension pour plonger les données dans un espace latent de dimension fixe et égale à soixante-quatre est mise en œuvre, et en sortie les similarités entre transformations sont scorées, les plus grands scores de similarité étant anormaux, tandis que les plus petits scores de similarité étant normaux.

- [0082] L'état des poids du ou des réseaux de neurones après l'entraînement est fixé et exporté en vue d'au moins une phase d'inférence P_2 subséquente.
- [0083] Ainsi, la phase d'apprentissage P_1 crée un « modèle » M comprenant plusieurs éléments:
- [0084] - les états des processeurs de variable (i.e. du module de prétraitement 14) permettant la traduction des données source en données adaptées au réseau de neurones et vice versa ;
- [0085] - les poids des couches du modèle M de réseau de neurones récurrents ;
- [0086] - les paramètres fixes tels que le nombre de transformations ou la température utilisée dans la fonction de perte contrastive ;
- [0087] - les valeurs minimale et maximale de la fonction de perte L rencontrées à l'entraînement (i.e. l'apprentissage).
- [0088] Comme précédemment indiqué, un tel modèle M est utilisé ensuite en inférence, c'est-à-dire que son application à un seul jeu de transactions (i.e. de données billettiques tabulaires de test) permet de fournir en sortie un score d'anomalie, le modèle M ayant appris ce qu'était une donnée anormale étant à la volée ainsi capable de dire d'un jeu de transactions (e.g. de vingt transaction) si ce jeu de transactions est anormal ou non.
- [0089] Ladite au moins une phase d'inférence P_2 comprend pour sa part tout d'abord une étape 40 d'application dudit prétraitement (noté P_I pour prétraitement en inférence) et de ladite sélection de colonnes de données d'intérêt prédéterminées auxdites données billettiques tabulaires de test, ledit prétraitement et ladite sélection appliquées aux données billettiques tabulaires de test étant identique à celui appliqué en phase d'apprentissage P_1 aux données billettiques tabulaires d'entraînement (comme détaillé par la suite en relation avec la [Fig.3] en termes de traitement de variable par chaque processeur de variable (i.e. du module de prétraitement 14 ou 23), puis de codage par un encodeur de position).
- [0090] De telles étapes 36 et 40 de prétraitement et sélection en apprentissage, comme de prétraitement et sélection en inférence sont davantage détaillées par la suite en relation avec la [Fig.3].
- [0091] Puis, ladite au moins une phase d'inférence P_2 comprend une étape 42 d'obtention d'un score S correspondant à un niveau d'anomalie desdites données billettiques ta-

bulaires de test par application en inférence dudit modèle M auxdites données billettiques de test optimisées par ledit prétraitement 40.

- [0092] En complément optionnel, l'obtention 42 dudit score S en phase d'inférence P_2 comprend une normalisation en utilisant les valeurs minimales et maximales de la fonction de perte contrastive rencontrées en phase d'apprentissage P_1 .
- [0093] Autrement dit, le score S , selon ce complément optionnel, est normalisé par les valeurs minimales et maximales de la fonction de perte rencontrées à l'entraînement (i.e. l'apprentissage) de telle sorte que, par exemple, un score normalisé supérieur à cent corresponde à une série de transactions qui n'a pas été vue auparavant par le réseau et donc très anormale.
- [0094] Un tel score S obtenu en inférence est un score d'anomalie indiquant seulement à l'opérateur si le schéma de transaction analysé est normal ou non. Un tel score S ne comprend aucune indication relative à l'origine de l'anomalie; et il est donc difficile de s'assurer s'il s'agit d'une fraude, car face aux données d'un réseau de transport en constante évolution, il est attendu de voir le réseau de neurone découvrir des schémas anormaux très complexes, si bien que la qualification des résultats du réseau de neurone nécessite une plus grande transparence.
- [0095] Pour répondre à ce besoin, de manière optionnelle complémentaire, la phase d'inférence P_2 comprend en outre une étape 44 d'explication, mise en œuvre par le module d'explicabilité 28, attribuant des poids aux colonnes sélectionnées lors de l'étape de prétraitement, lesdits poids étant propres à expliquer au moins en partie ledit score obtenu.
- [0096] A titre d'exemple, pour mettre en œuvre une telle étape d'explication 44, le module d'explicabilité 28 utilise les résultats, décrits par Sundararajan et al. dans l'article intitulé « Axiomatic Attributions for Deep Networks » de 2017, et introduisant les gradients intégrés; pour attribuer des poids aux colonnes du score d'anomalie final, une colonne étant pour ce faire encodée sur plusieurs "sous-colonnes", le poids étant propre à être calculé en utilisant une intégrale le long du chemin du modèle de réseau de neurones.
- [0097] Via cette méthode la colonne correspondant au type de transaction présente par exemple un poids positif de 0,091215, la colonne correspondant au type de titre (i.e. support billettique, tel qu'à Paris, la carte Navigo® mensuelle, ou encore la carte Navigo® liberté +, le passe Navigo® easy etc.) présente un poids négatif - 0,14268, et la colonne correspondant au montant de la transaction présente un poids positif de 1,788538, les poids attribués à chaque colonne étant relatifs. Selon cet exemple, au vu des poids précités, le montant des transactions étudiées augmente la valeur de la fonction de perte et donc contribue positivement au score d'anomalie, si bien que l'opérateur peut creuser cette piste en étudiant en premier lieu les transactions dont les

montants sont anormaux.

- [0098] Selon un autre exemple, un score d'anomalie égal à cent-quinze est obtenu en considérant neuf colonnes sélectionnées en entrée du réseau de neurones avec vingt lignes correspondant aux dernières transactions, dont une colonne correspondant aux entrées/sorties, une entrée étant représentée par le chiffre 1 et une sortie représentée par le chiffre 2. Le poids le plus haut de contribution au score, à savoir à hauteur de 64,9%, est justement celui associé à cette colonne d'entrées/sorties notamment du fait que trois transactions successives du même utilisateur (ou associées à un unique identifiant d'un titre de transport en commun ou d'un moyen de paiement grâce auquel cet utilisateur voyage) sont représentées par une suite de trois un, ce qui correspond à trois entrées successives alors que les entrées devraient alterner avec les sorties. L'opérateur au vu de cette explication vérifie alors la numérotation des transactions associées, et cette vérification montre que les numéros de transaction se suivent et qu'aucune transaction n'est manquante, si bien que l'opérateur en déduit que l'anomalie associée au score égal à cent-quinze n'est pas une fraude mais probablement une anomalie liée au réseau, et correspond à des sorties qui ne sont pas correctement enregistrées comme telle si bien qu'il est nécessaire de mettre en œuvre une opération de maintenance.
- [0099] Plus généralement, un poids au signe positif indique que la contribution de la colonne considérée fait tendre le score vers de plus grandes valeurs; tandis qu'un poids au signe négatif fait tendre le score vers de plus petites valeurs. L'ampleur de la contribution est corrélée à la valeur du poids (zéro signifiant une petite contribution).
- [0100] Selon une étape 46, mise en œuvre mentalement par un opérateur dudit système électronique 10 de détection d'anomalies dans les données billettiques de transport en commun, une décision est prise pour caractériser les données billettiques tabulaires de test comme des données porteuses (i.e. représentatives) d'anomalie(s) ou comme des données normales.
- [0101] La [Fig.3] illustre en détail les étapes 36 (comme 40 en inférence) de prétraitement de données et sélection de colonnes de données d'intérêt prédéterminées au sein desdites données billettiques tabulaires d'entraînement en apprentissage (respectivement de test en inférence).
- [0102] Plus précisément, selon le mode de réalisation de la [Fig.3], le prétraitement 36, 40 et la sélection de colonnes de données d'intérêt prédéterminées comprend tout d'abord une phase 48 de définition DEF permettant de définir les différentes variables à utiliser ainsi que les paramètres choisis pour l'entraînement du ou des réseaux de neurones utilisés par la suite lors de l'apprentissage profond 38.
- [0103] Cette phase 48 de définition comprend notamment une sous-étape 50 de sélection, pour chaque utilisateur de transport en commun (i.e. voyageur ou par extension associé à un unique identifiant d'un titre de transport en commun ou d'un moyen de paiement

grâce auquel cette personne voyage), d'un nombre prédéterminé L_T de dernières transactions billettiques et des données billettiques tabulaires associées. Par exemple, au regard des habitudes de transport des usagers, le nombre prédéterminé L_T de dernières transactions billettiques est fixé aux vingt dernières transactions de chaque utilisateur de transport en commun (i.e. voyageur).

[0104] De plus, cette phase 48 de définition comprend, au sein desdites données billettiques tabulaires associées au nombre prédéterminé de dernières transactions, une sous-étape 52 de sélection des colonnes, notées par exemple $C_{D,T}$, de données correspondant au type de transaction. Par « type de transaction », on entend ici la nature de chaque transaction, telle que la transaction de type voyage, de type rechargement, etc. Autrement dit, la sous-étape 52 consiste à privilégier (i.e. sélectionner) les colonnes des données billettiques tabulaires contenant des détails sur les transactions.

[0105] De plus, cette phase 48 de définition comprend, au sein desdites données billettiques tabulaires associées au nombre prédéterminé de dernières transactions, une sous-étape 54 de sélection (i.e. d'ajout aux données sélectionnées pour apprentissage profond) des colonnes, notées par exemple $A_{V,Q}$, portant sur la valeur numérique des transactions de l'utilisateur (i.e. du voyageur ou par extension associé à un unique identifiant d'un titre de transport en commun ou d'un moyen de paiement grâce auquel cette personne voyage) considéré. Par « valeur numérique de transaction », on entend ici une variable quantitative telle que la quantité de monnaie locale au rechargement, la quantité de monnaie locale payée à chaque transaction, cette variable étant relative, notamment négative en cas de paiement ou positive en cas de chargement ou de rechargement du titre de transport de l'utilisateur considéré.

[0106] De plus, cette phase 48 de définition comprend, au sein desdites données billettiques tabulaires associées au nombre prédéterminé de dernières transactions, une sous-étape 54 de sélection (i.e. d'ajout aux données sélectionnées pour apprentissage profond) des colonnes, notées par exemple $C_{D,V}$, portant sur le contexte billettique de transaction. Par « contexte billettique de transaction », on entend les colonnes qui concernent la description du voyage associé à la transaction, telles que les incréments de transaction de titres, les types d'usage tels que les entrées/sorties de gare ou dans un espace ayant une fonction particulière du point de vue de la billettique ou des infrastructures associées (par exemple : un véhicule ou une partie d'un véhicule, une zone particulière d'une infrastructure de transport ou de stationnement).

[0107] En complément optionnel, l'étape 36 de prétraitement et de sélection de colonnes de données d'intérêt prédéterminées comprend en outre une sous-étape 58 de formatage (notée FORMAT) desdites données billettiques sélectionnées (par exemple au format Pandas).

[0108] En effet, chaque table de transaction (i.e. donnée billettique tabulaire) à

l'entraînement (i.e. l'apprentissage) est propre à se présenter selon plusieurs formats, notamment ascii, lisible ou non par une machine, etc.

- [0109] L'apprentissage profond 38 selon la présente invention nécessite un format de donnée d'entrée du réseau de neurone systématiquement lisible par une machine (de l'anglais « *machine-readable* »), et la sous-étape 58 de formatage au format pré-déterminé associé au réseau de neurones considéré permet d'assurer une telle lecture systématique.
- [0110] En complément optionnel, l'étape 36 de prétraitement et de sélection de colonnes de données d'intérêt prédéterminées comprend en outre une sous-étape 60 de tri T des transactions dans l'ordre chronologique suivi d'un groupement desdites transactions par utilisateur (ou par extension associé à un unique identifiant d'un titre de transport en commun ou d'un moyen de paiement grâce auquel cette personne voyage), chaque utilisateur étant identifié par un identifiant unique, et disponible à l'origine dans les tables de transactions billettiques en entrée dudit prétraitement 36, un tel identifiant unique étant indiqué par exemple dans une colonne par exemple intitulée MediaSerial-NumberID pour les données billettiques tabulaires du système billettique selon la Demanderesse Revenue Collection Systems
- [0111] De plus, selon ce complément optionnel, le tri et groupement 60 sont suivis d'une sous-étape 62 de filtrage F des utilisateurs afin de ne conserver que les données billettiques associées aux utilisateurs ayant un nombre de transactions supérieur ou égal audit nombre prédéterminé L_T de dernières transactions billettiques. Autrement dit, un tel filtrage revient à ne conserver les utilisateurs (i.e. les voyageurs) qui présentent plus de L_T transactions par exemple vingt transactions, les utilisateurs associés à un nombre de transactions moindre correspondant généralement à des utilisateurs non locaux (i.e. voyageurs non locaux).
- [0112] Selon un autre complément optionnel, l'étape 36 de prétraitement et de sélection de colonnes de données d'intérêt prédéterminées comprend en outre une phase supplémentaire 64 de transformation(s) de donnée(s) afin que chaque donnée soit adaptée à l'entrée dans un réseau de neurones. Une telle phase 64 de transformation(s) supplémentaire(s) comprend une sous-étape 66 de typage suivie d'une sous-étape 68 de séparation des données billettiques tabulaires sélectionnées en deux types de variables comprenant des variables catégorielles propres à décrire qualitativement un état et/ou un élément billettique et des variables numériques propres à décrire quantitativement un état et/ou un élément billettique.
- [0113] Le typage 66 des variables de colonnes de données billettiques tabulaires permet de s'assurer que celles-ci ont un format tel qu'attendu en entrée du réseau de neurones subséquent, puis séparées au cours de l'étape 68 selon les deux types précités à savoir d'une part les variables catégorielles correspondant aux colonnes contenant peu de

valeurs différentes et décrivant un état du système billettique, et les variables numériques pouvant prendre une valeur a priori arbitraire sur l'intervalle des valeurs possibles; une transaction pouvant être positive, ou négative mais a priori entière.

- [0114] Autrement dit, le module 14 de prétraitement avant apprentissage et le module 23 de prétraitement avant inférence configurés pour mettre en œuvre respectivement les prétraitements 36 en apprentissage et 40 en inférence, sont des processeurs de variables propres à coder les variables catégorielles par exemple selon un codage connu par le terme anglais One-Hot-Encoding où une suite de variables ABBCCC est par exemple encodée au moyen du tableau suivant de valeur binaire :

A	B	C
1	0	0
0	1	0
0	1	0
0	0	1
0	0	1
0	0	1

- [0115] Les variables numériques sont codées en utilisant un outil de mise à l'échelle standard (de l'anglais *StandardScaler*) où les valeurs sont normalisées en moyenne et en variance par rapport à la statistique de l'ensemble de la distribution numérique.
- [0116] Les résultats fournis par de tels processeurs de variables sont stockés sous forme de dictionnaire qui permettent de transformer de façon transparente et unique chaque variable de la version brute D_T (données tabulaire d'entrée) à la version encodée D_O (données optimisées d'entrée de réseau(x) de neurone(s)) et vice versa.
- [0117] En outre, en complément optionnel, le codage proposé selon la présente invention est propre à ajouter aux variables un double encodage, à savoir en premier lieu de la position de chaque transaction dans la séquence des L_T dernières transactions, notamment tel que L_T=20, et en deuxième lieu un encodage de ces positions (i.e. des positions de chaque transaction) par un encodeur de position dans un espace latent de plus faible dimension pour gagner en espace de traitement de donnée (i.e. en anglais de *data processing*).
- [0118] Autrement dit, à l'issue du prétraitement 36, on obtient en entrée du réseau de neurones un dictionnaire de tenseurs (i.e. des tableaux numériques) contenant des valeurs encodées liées aux valeurs brutes des données billettiques sources D_T fournie à l'origine en entrée de la phase d'apprentissage ou de la phase d'inférence.
- [0119] En d'autres termes, le prétraitement 36 ou 40 mis en œuvre selon la présente

invention permet d'atteindre le niveau d'abstraction nécessaire en entrée du ou des réseau(x) de neurones utilisés par la suite en phase d'apprentissage comme en phase d'inférence, les réseaux de neurones étant généralement optimisés pour des valeurs numériques proches de zéro.

- [0120] Ces valeurs de données optimisées par ledit prétraitement 36 ou 40 sont utilisées par au moins un réseau de neurones lors de la phase d'apprentissage P_1 afin d'optimiser son propre comportement pour créer un modèle à utiliser ensuite sur des données tabulaires à tester. Autrement dit, un tel modèle, léger, une fois confronté à un jeu de transactions (i.e. des données billettiques tabulaires de test), est propre à fournir un score S qui correspond au niveau d'anomalie des données billettiques tabulaires de test.
- [0121] L'homme du métier comprendra que l'invention ne se limite pas aux modes de réalisation décrits, ni aux exemples particuliers de la description, les modes de réalisation et les variantes mentionnées ci-dessus étant propres à être combinés entre eux pour générer de nouveaux modes de réalisation de l'invention.
- [0122] L'homme du métier comprendra que l'invention peut être utilisée pour traiter des données associées à une application de billettique pour les transports en commun ainsi qu'à des applications de billettiques pouvant être combinées avec des transports en commun ; elle peut par exemple être appliquée à un système billettique qui combinerait des données liées à des transports en communs avec des données de type transaction de stationnement ou liées à l'utilisation de moyens de transports qui peuvent être utilisés en complément de transports en commun (par exemple : utilisation de moyens de transports individuels : vélos ou trottinettes en libre-service, ...)
- [0123] La présente invention propose ainsi un prétraitement spécifique au format de données tabulaires utilisé dans les système billettiques de transport en commun permettant d'augmenter la vitesse de traitement de données, et utilise des réseaux de neurones, notamment des réseaux de neurones convolutifs, avec une fonction de perte contrastive pour analyser les transformations (i.e. vues) de données billettiques et catégoriser les données anormales comme telles de façon non supervisée, en utilisant une échelle de valeurs, la plus basse étant normale et la plus haute étant anormale, le réseau de neurones entraîné étant propre à fournir une probabilité que la transaction immédiatement reçue soit normale ou non.
- [0124] La présente invention permet par exemple d'obtenir à l'issue d'une phase d'entraînement en cinquante étapes en considérant neuf colonnes de données billettiques tabulaires collectées sur une période de trois mois sélectionnées et pré-traitées selon la présente invention, un modèle en dix-huit heures sur une unité de traitement graphique GPU (de l'anglais Graphics Processing Unit), telle que la GPU NVIDIA® Tesla P100 12 Gb, sachant qu'en phase d'entraînement l'objectif est

d'obtenir un modèle performant sans contrainte de temps, l'entraînement étant « hors ligne » (de l'anglais off-line). En revanche, en inférence (i.e. en exploitation), la présente invention permet au minimum de scorer cent transactions par seconde pour s'adapter à un réseau de transport de taille moyenne, et jusqu'à quatre-cents transactions par seconde, en considérant que cent millisecondes sont nécessaires pour charger le modèle issu de la phase d'apprentissage et pour le prétraitement 40, et que cent millisecondes sont au minimum nécessaires pour effectuer l'inférence d'une transaction, ce traitement d'inférence étant propre à être parallélisé pour un nombre de transactions allant jusqu'à 20 000 transactions.

[0125] La présente invention propose ainsi une application au comportement utilisateur et permet ainsi, d'améliorer la réactivité et l'efficacité des opérateurs face à des tentatives de fraudes sophistiquées, tout en faisant face, de façon évolutive, à des mécanismes de fraude complexes et encore inconnus des opérateurs.

Revendications

[Revendication 1]

Procédé (34) de détection d'anomalies dans les données billettiques de transport en commun, le procédé étant propre à être mis en oeuvre par un système électronique de détection d'anomalies dans les données billettiques de transport en commun, le procédé comprenant au moins une phase d'apprentissage (P_1) et au moins une phase d'inférence (P_2), ladite au moins une phase d'apprentissage (P_1) comprenant les étapes suivantes, appliquées à des données billettiques tabulaires

d'entraînement :

- prétraitement et sélection (36) de colonnes de données d'intérêt prédéterminées au sein desdites données billettiques tabulaires

d'entraînement ;

- apprentissage profond (38) d'au moins un réseau de neurones en utilisant les données billettiques d'entraînement optimisées par ledit prétraitement et obtention d'un modèle dudit au moins un réseau de neurones configuré pour détecter une anomalie de données billettiques ; ladite au moins une phase d'inférence (P_2) comprenant les étapes suivantes, appliquées à des données billettiques tabulaires de test ;

- application (40) dudit prétraitement et de ladite sélection de colonnes de données d'intérêt prédéterminées auxdites données billettiques tabulaires de test ;

- obtention (42) d'un score correspondant à un niveau d'anomalie desdites données billettiques tabulaires de test par application dudit modèle auxdites données billettiques de test optimisées par ledit prétraitement.

[Revendication 2]

Procédé (34) selon la revendication 1, dans lequel le prétraitement et la sélection (36) de colonnes de données d'intérêt prédéterminées comprennent :

- la sélection (50), pour chaque utilisateur de transport en commun, ou par extension pour chaque unique identifiant d'un titre de transport en commun ou d'un moyen de paiement grâce auquel cette personne voyage, d'un nombre prédéterminé de dernières transactions billettiques et des données billettiques tabulaires associées ;

- au sein desdites données billettiques tabulaires associées au nombre prédéterminé de dernières transactions, la sélection (52, 54, 56) des colonnes de données correspondant au :

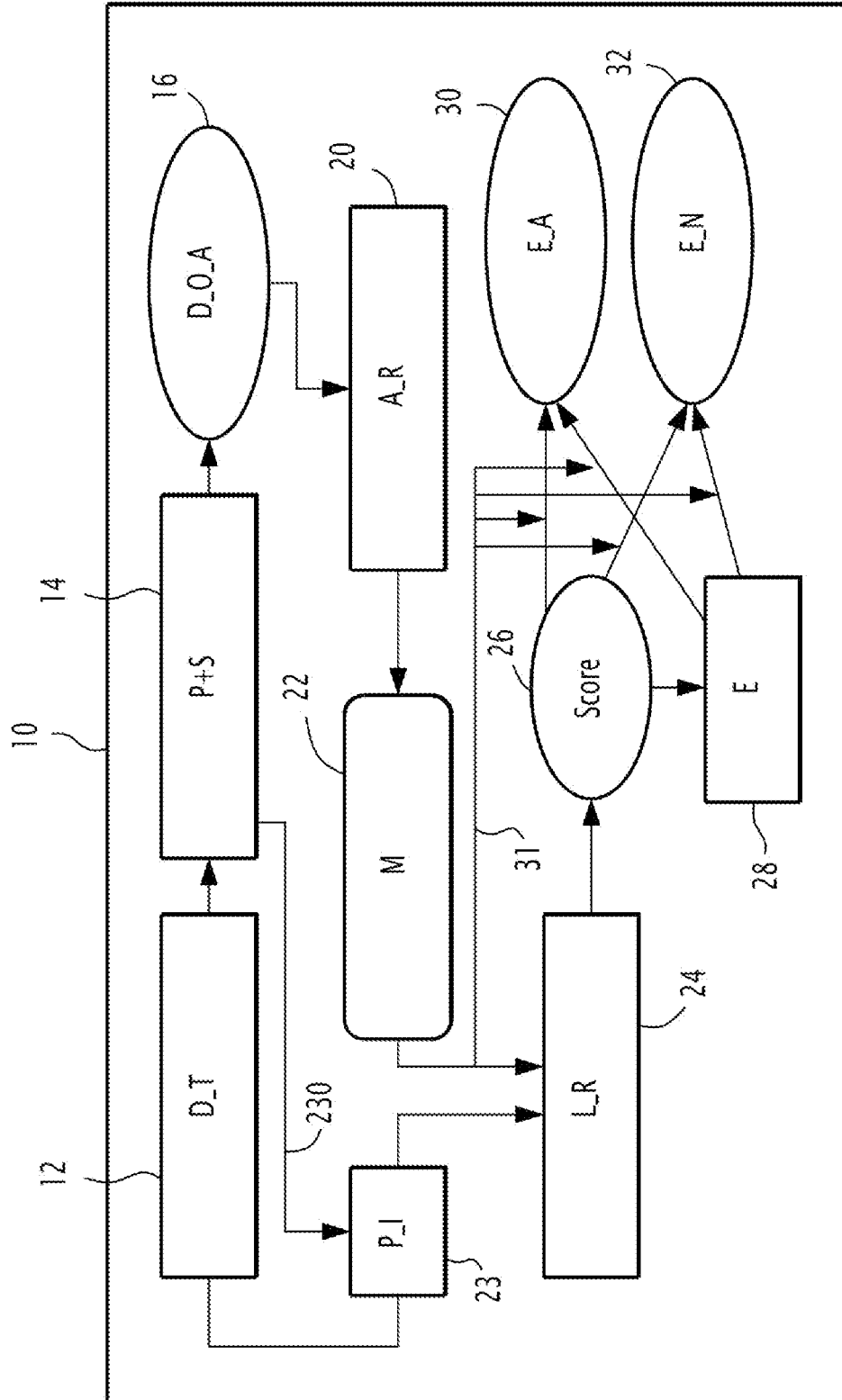
- type de transaction ;

- la valeur numérique de transaction ;
- le contexte billettique de transaction.

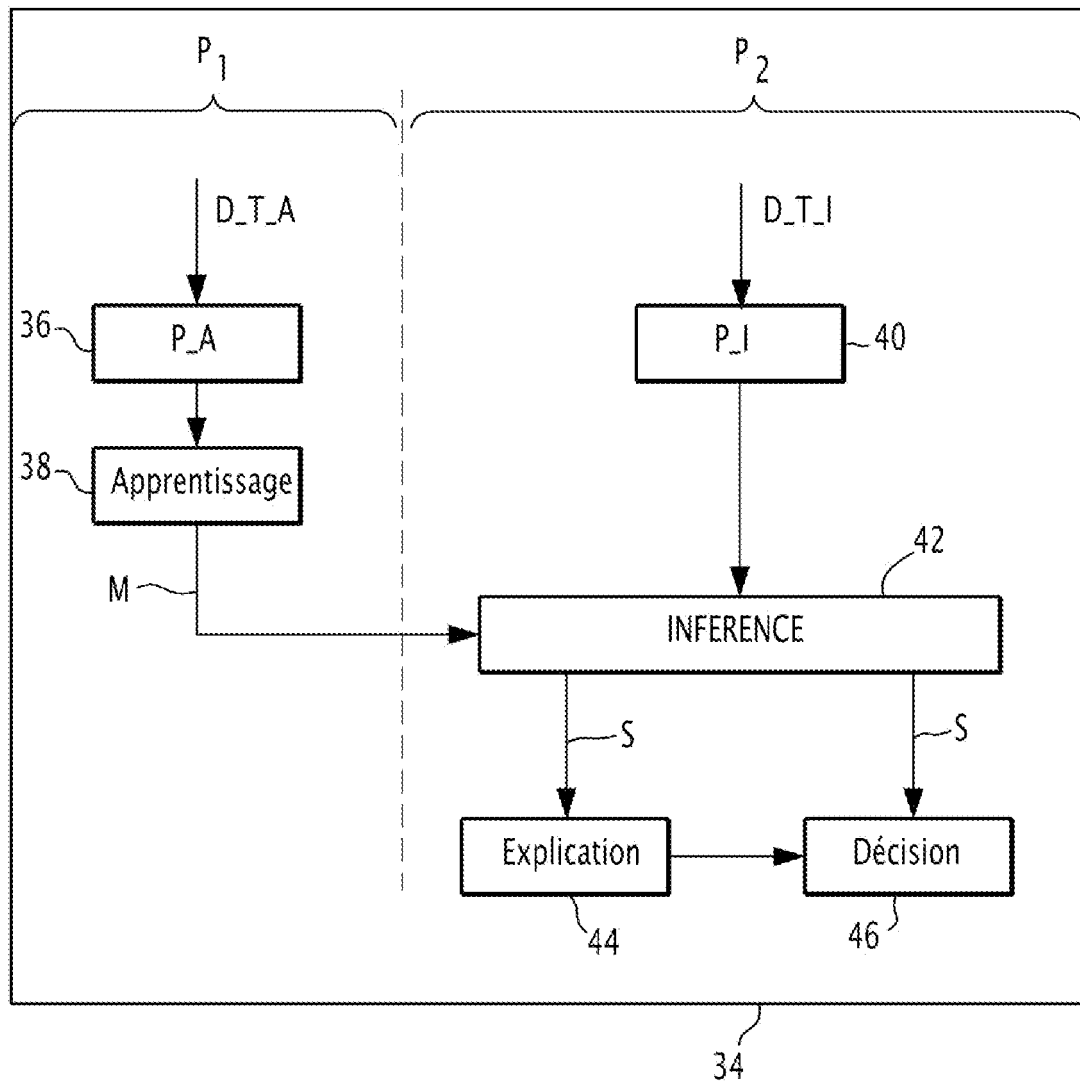
- [Revendication 3] Procédé (34) selon la revendication 1 ou 2, dans lequel le prétraitement et la sélection (36) de colonnes de données d'intérêt prédéterminées comprennent en outre un formatage (58) desdites données billettiques sélectionnées selon un format prédéterminé.
- [Revendication 4] Procédé (34) selon la revendication 2 ou 3, dans lequel le prétraitement et la sélection (36) de colonnes de données d'intérêt prédéterminées comprennent en outre un tri (60) des transactions dans l'ordre chronologique suivi d'un groupement desdites transactions par utilisateur, chaque utilisateur étant identifié par un identifiant unique, puis un filtrage des utilisateurs afin de ne conserver que les données billettiques associées aux utilisateurs ayant un nombre de transactions supérieur ou égal audit nombre prédéterminé de dernières transactions billettiques.
- [Revendication 5] Procédé (34) selon l'une quelconque des revendications précédentes, dans lequel le prétraitement et la sélection (36) de colonnes de données d'intérêt prédéterminées comprennent en outre un typage (66) puis une séparation (68) des données billettiques tabulaires sélectionnées en deux types de variables comprenant des variables catégorielles propres à décrire qualitativement un état et/ou un élément billettique et des variables numériques propres à décrire quantitativement un état et/ou un élément billettique.
- [Revendication 6] Procédé (34) selon l'une quelconque des revendications précédentes, dans lequel ledit apprentissage profond (38) utilise une fonction de perte contrastive et dans lequel l'obtention (42) dudit score en phase d'inférence comprend une normalisation en utilisant les valeurs minimales et maximales de la fonction de perte contrastive rencontrées en phase d'apprentissage.
- [Revendication 7] Procédé (34) selon l'une quelconque des revendications précédentes, dans lequel ledit apprentissage profond (38) comprend :
- la détermination, au moyen dudit au moins un réseau de neurones, d'une pluralité de vues des données billettiques issues dudit prétraitement et de ladite sélection,
 - la réduction dimensionnelle de chacune des vues de ladite pluralité de vues fournissant des vues dans un espace latent de dimension réduite ;
 - le calcul de propriétés géométriques prédéterminées dans ledit espace latent de plus petite dimension obtenue par ladite réduction dimensionnelle.

- [Revendication 8] Procédé (34) selon l'une quelconque des revendications précédentes, dans lequel ladite au moins une phase d'inférence comprend en outre une étape (44) d'explication attribuant des poids aux colonnes sélectionnées lors de l'étape de prétraitement, lesdits poids étant propres à expliquer au moins en partie ledit score obtenu.
- [Revendication 9] Programme d'ordinateur comportant des instructions logicielles qui lorsqu'elles sont exécutées par un ordinateur, mettent en œuvre un procédé de détection d'anomalies dans les données billettiques de transport en commun selon l'une quelconque des revendications précédentes.
- [Revendication 10] Système électronique (10) de détection d'anomalies dans les données billettiques de transport en commun, le système électronique comprenant une unité d'apprentissage et une unité d'inférence, l'unité d'apprentissage comprenant, pour application à des données billettiques tabulaires d'entraînement :
- un module (14) de prétraitement configuré pour prétraiter et sélectionner des colonnes de données d'intérêt prédéterminées au sein desdites données billettiques tabulaires d'entraînement ;
 - un module (20) d'apprentissage profond configuré pour mettre en œuvre un apprentissage profond d'au moins un réseau de neurones en utilisant les données billettiques d'entraînement optimisées par ledit prétraitement et pour obtenir un modèle (22) dudit au moins un réseau de neurones propre à détecter une anomalie de données billettiques ;
- ladite unité d'inférence comprenant, pour application à des données billettiques tabulaires de test :
- un module (23) d'application configuré pour appliquer ledit prétraitement et ladite sélection de colonnes de données d'intérêt prédéterminées auxdites données billettiques tabulaires de test ;
 - un module (24) d'obtention d'un score (26) correspondant à un niveau d'anomalie desdites données billettiques tabulaires de test, ledit module d'obtention (24) étant configuré pour obtenir ledit score par application dudit modèle (22) auxdites données billettiques de test optimisées par ledit prétraitement (230).

[Fig. 1]



[Fig. 2]



[Fig. 3]

