

①⑨ RÉPUBLIQUE FRANÇAISE  
—  
**INSTITUT NATIONAL  
DE LA PROPRIÉTÉ INDUSTRIELLE**  
—  
COURBEVOIE  
—

①① N° de publication :

**3 071 085**

(à n'utiliser que pour les  
commandes de reproduction)

②① N° d'enregistrement national :

**17 58514**

⑤① Int Cl<sup>8</sup> : **G 06 Q 30/02** (2020.12), G 06 F 16/951, G 06 F 17/  
40, G 06 Q 50/14

①②

## BREVET D'INVENTION

**B1**

⑤④ UN PROCEDE ET UN SYSTEME D'APPRENTISSAGE AUTOMATIQUE POUR PREDIRE LES  
INTERACTIONS D'UN UTILISATEUR EN LIGNE.

②② Date de dépôt : 14.09.17.

③③ Priorité :

④③ Date de mise à la disposition du public  
de la demande : 15.03.19 Bulletin 19/11.

④⑤ Date de la mise à disposition du public du  
brevet d'invention : 04.02.22 Bulletin 22/05.

⑤⑥ Liste des documents cités dans le rapport de  
recherche :

*Se reporter à la fin du présent fascicule*

⑥⑥ Références à d'autres documents nationaux  
apparentés :

○ Demande(s) d'extension :

⑦① Demandeur(s) : *Amadeus S.A.S. — FR.*

⑦② Inventeur(s) : *ACUNA AGOST RODRIGO et  
RENAUDIE DAVID.*

⑦③ Titulaire(s) : *Amadeus S.A.S..*

⑦④ Mandataire(s) : *SAMSON & PARTNER  
PATENTANWALTE MBB.*

**FR 3 071 085 - B1**



# UN PROCÉDÉ ET UN SYSTÈME D'APPRENTISSAGE AUTOMATIQUE POUR PRÉDIRE LES INTERACTIONS D'UN UTILISATEUR EN LIGNE

## DOMAINE DE L'INVENTION

[0001] La présente invention concerne l'application de modèles d'apprentissage automatique pour prédire le comportement des utilisateurs en ligne. En particulier, des modes de réalisation de l'invention prédisent la probabilité d'une interaction d'un utilisateur avec des éléments de contenu en ligne sur la base du comportement agrégé des utilisateurs précédents dans des contextes similaires. L'invention peut s'appliquer aux systèmes de publicité en ligne, par exemple pour déterminer la pertinence de la d'une publicité présentée à l'utilisateur, p. ex. via une page Web ou dans une application mobile.

## CONTEXTE DE L'INVENTION

[0002] La publicité en ligne (p. ex. basée sur le Web, les applications intégrées ou mobiles) diffère de la publicité des médias traditionnels en matière de degré de ciblage personnalisé du public. Par exemple, alors que la publicité des supports de diffusion tels que la publicité diffusée à la télévision vise à atteindre une cible démographique définie par des caractéristiques élargies telles que la tranche d'âge, le statut socio-économique et/ou des intérêts généraux, la publicité en ligne cible des individus ayant un intérêt particulier pour le produit, le service ou l'information qui est présentée.

[0003] La technologie de ciblage hautement personnalisé d'un public a conduit au développement de modèles commerciaux qui sont spécifiques à la publicité en ligne. Par exemple, il est maintenant courant pour les sites Web qui fournissent des actualités, des informations agrégées et d'autres contenus d'intérêt à des utilisateurs particuliers, d'héberger des publicités de tiers comme moyen de générer des revenus. Les annonceurs dont les annonces apparaissent sur ces sites Web peuvent payer l'opérateur sur la base des opportunités de

consultation ou d'impression (couramment mesurées comme 'coût par milliers d'impressions', autrement dit CPM), sur la base du coût par clic (CPC) ou selon d'autres mesures de performance. La sélection actuelle d'une publicité devant être placée sur une page Web présentée à un utilisateur individuel peut être basée, au moins en partie, sur un processus d'enchère par lequel un annonceur qui est disposé à payer plus pour un CPM, CPC ou autre mesure de coût a de meilleures chances que son annonce soit présentée à l'utilisateur.

[0004] Selon un modèle courant, le processus d'enchère est facilité par une 'plateforme de bourse publicitaire'. Une bourse publicitaire est une plateforme technologique qui implémente un marché numérique permettant aux annonceurs et aux diffuseurs des sites Web et autres contenus en ligne d'acheter et de vendre des encarts publicitaires, souvent par le biais d'enchères en temps réel. Des plateformes de bourses publicitaires bien connues incluent DoubleClick™ (propriété de Google™), AppNexus™, Microsoft™ Ad Exchange™, et OpenX™.

[0005] Une bourse publicitaire maintient une 'réserve' d'encarts publicitaires. Les annonceurs fournissent leur encart publicitaire à la réserve, p. ex. les encarts publicitaires disponibles, intégrés dans les pages Web qui sont distribués aux utilisateurs. Les acheteurs peuvent ensuite enchérir pour les encarts qu'ils souhaitent acheter. Les décisions d'enchérir sont souvent faites en temps réel sur la base d'informations telles que le comportement antérieur de l'utilisateur auquel une publicité a été présentée, le moment de la journée, le type de dispositif, la position de la publicité et ainsi de suite. En pratique ces décisions d'enchère doivent elles-mêmes être faites très rapidement, p. ex. de l'ordre de quelques dixièmes de millisecondes, en utilisant des plateformes technologiques couramment connues sous le nom de plateformes de demande (DSPs). Vu que le coût de l'achat d'impressions sur la bourse publicitaire est réel pour l'annonceur, la performance des technologies et des algorithmes déployés par une DSP pour évaluer la 'valeur' potentielle d'un utilisateur dans le but de décider d'enchérir peut avoir un impact commercial substantiel.

[0006] À titre d'exemple, un taux moyen par clic (CTR) pour l'affichage des publicités sur le Web est environ 0,05 % (c.-à-d. cinq clics pour 10 000 impressions. Dans lequel, les encarts publicitaires achetés via une bourse d'échanges publicitaires représentent un coût (c.-à-d. le prix payé après une enchère gagnante). Cependant, lorsque le modèle CPC est utilisé, seul 0,05 % des impressions en moyenne donnent lieu à la génération d'un revenu. Donc un CTR faible donne lieu à une utilisation inefficace des ressources technologiques (telles que les ressources de traitement de la DSP) et un coût plus élevé pour les annonceurs gagnants (puisque un plus petit nombre de 'clics' doit couvrir le coût de toutes les impressions). En outre, un faible CTR est indicatif d'un manque d'adéquation ou de pertinence de la publicité affichée aux utilisateurs en ligne.

[0007] Il est donc extrêmement souhaitable de déployer des technologies sur les plateformes DSP qui engendrent un CTR plus élevé. Ces technologies sont bénéfiques pour les opérateurs DSP en améliorant l'utilisation des ressources techniques et en augmentant les revenus générés à partir des interactions de l'utilisateur avec le contenu publicitaire. Les annonceurs bénéficient aussi d'un placement plus fructueux et plus efficace des publicités, aboutissant à des clics plus nombreux et une diffusion à un public plus large de clients potentiellement intéressés. Les utilisateurs en ligne sont aussi gagnants puisque un CTR plus élevé est généralement atteint par le placement de publicités qui sont plus appropriées et pertinentes pour les utilisateurs.

[0008] Une approche courante pour augmenter le CTR consiste à faire un suivi pour inférer les intérêts de l'utilisateur et pour placer le contenu publicitaire qui est pertinent par rapport à leurs intérêts. Par exemple, les témoins (cookies) du navigateur et autres technologies de suivi peuvent être utilisés pour amasser des informations concernant par exemple des sites Web visités par l'utilisateur, des sujets consultés sur les sites de réseaux sociaux et/ou des recherches effectuées en ligne par l'utilisateur. Ces informations peuvent être traitées pour identifier le contenu publicitaire dont on peut inférer qu'il est pertinent pour l'utilisateur. Cependant, la prédiction de la probabilité que l'utilisateur ait une interaction avec ce contenu est plus complexe. Comme souligné ci-dessus, la

moyenne CTR est faible et les utilisateurs individuels en ligne interagissent rarement avec le contenu publicitaire même s'il a été ciblé de façon effective sur la base des intérêts de l'utilisateur. Par conséquent il y a très peu de données directes qui peuvent être utilisées pour prédire la probabilité que les utilisateurs interagissent avec un contenu publicitaire sur la base du suivi d'un utilisateur individuel.

[0009] Par conséquent, il y a clairement un besoin de technologies, de procédés et de systèmes, mis en œuvre informatiquement en temps réel, qui peuvent être déployés dans les DSPs et qui sont capables de prendre des décisions améliorées concernant la sélection des publicités et de placer une enchère sur les plateformes de bourses publicitaires. Il serait en particulier désirable d'apporter des procédés et des systèmes améliorés pour prédire la probabilité qu'un utilisateur en ligne interagisse avec le contenu publicitaire. Afin de satisfaire les exigences techniques et d'assurer que le temps de chargement d'une page ne soit pas inutilement accru par la sélection et le processus d'enchère, il est nécessaire que cette prédiction soit faite en quelques dixièmes de millisecondes. La présente invention a pour objectif de répondre à ces besoins.

## RÉSUMÉ DE L'INVENTION

[0010] Dans un aspect, la présente invention fournit un procédé mis en œuvre informatiquement comprenant :

l'accès à un entrepôt de données en ligne pour récupérer des enregistrements concernant des événements de placement de contenu et des enregistrements concernant des événements d'interaction d'un utilisateur, dans lequel les événements de placement et d'interaction surviennent au cours d'un laps de temps défini ;

l'appariement des enregistrements récupérés d'événements de placement de contenu avec des enregistrements récupérés d'événements d'interaction pour

générer un ensemble de données appariées incluant un ensemble de valeurs brutes de caractéristiques dérivées d'un événement de placement de contenu avec une balise d'événements d'interaction indiquant si oui ou non un événement d'interaction correspondant avec l'événement de placement de contenu est survenu ;

le calcul, à partir des valeurs de caractéristiques brutes, d'un ensemble correspondant de vecteurs de caractéristiques de formation enrichies ;

la formation d'un modèle d'apprentissage automatique en utilisant les vecteurs de caractéristiques de formation et les balises d'événements d'interaction correspondantes ;

la réception, au niveau d'un processeur configuré pour exécuter le modèle d'apprentissage automatique, d'informations concernant l'encart de placement de contenu en ligne et des informations concernant un utilisateur auquel le contenu à l'intérieur de l'encart de placement de contenu en ligne sera affiché ;

le calcul, par le processeur, d'un vecteur de caractéristiques d'estimation, basé sur un élément de contenu sélectionné pour le placement à l'intérieur de l'encart de placement de contenu en ligne, d'informations concernant l'utilisateur et d'informations concernant l'encart de placement de contenu en ligne ;

la détermination, par le processeur exécutant le modèle d'apprentissage automatique, d'une estimation de la probabilité d'une interaction de l'utilisateur avec l'élément de contenu sélectionné, basée sur le vecteur de caractéristiques d'estimation enrichies.

[0011] De façon avantageuse, les modes de réalisation de l'invention utilisent l'appariement d'événements agrégés de placement de contenu avec des événements agrégés d'interactions d'utilisateurs pour soutenir la prédiction des interactions d'un utilisateur en ligne avec du contenu en utilisant un modèle d'apprentissage automatique. Les éléments de contenu peuvent être par exemple des publicités en ligne comprenant une ou plusieurs offres et les événements d'interaction de l'utilisateur peuvent comprendre des interactions telles que des clics sur des offres spécifiques dans les publicités. Par conséquent, dans ces modes de réalisation :

l'encart de placement de contenu en ligne est un encart publicitaire ;

les informations relatives à l'encart publicitaire et les informations relatives à l'utilisateur auquel le contenu à l'intérieur de l'encart publicitaire sera affiché sont reçues avec un message de demande d'enchère transmis par un serveur de bourse publicitaire ; et

l'élément de contenu comprend au moins une enchère pour un placement à l'intérieur d'un encart publicitaire.

[0012] Le procédé peut comprendre par ailleurs :

la transmission, par le processeur, à un serveur de bourse publicitaire, d'un message de réponse à une enchère en réponse au message de demande d'enchère ;

la réception par le processeur en provenance du serveur de bourse publicitaire, d'une notification d'enchère réussie ;

la mise à jour par le processeur d'un entrepôt de données en ligne avec des données d'événements de placement de contenu concernant le placement de l'élément de contenu

la réception par le processeur d'une notification d'interaction d'un utilisateur avec l'élément de contenu ; et

la mise à jour (226) de l'entrepôt de données en ligne (166) avec les données d'événements d'interaction d'utilisateur concernant l'interaction de l'utilisateur avec l'élément de contenu.

[0013] De cette façon, les modes de réalisation de l'invention prévoient une actualisation continue des enregistrements d'événements de placement de contenu et d'événements d'interaction d'utilisateur stockés, permettant de mettre à jour le modèle d'apprentissage automatique avec des informations actualisées concernant le comportement d'un utilisateur. Dans ce but, le procédé peut comprendre l'exécution récurrente des étapes pour accéder à l'entrepôt de données en ligne, l'appariement des enregistrements d'événements de placement de contenu récupérés avec les enregistrements d'événements d'interaction récupérés, le calcul de vecteurs de caractéristiques enrichies de formation et la formation du modèle d'apprentissage automatique.

[0014] Dans un autre aspect, l'invention fournit un appareil informatique qui implémente une plateforme du côté de la demande, l'appareil informatique comprenant :

un processeur ;

au moins un dispositif de mémoire accessible par le processeur ; et

une interface de communications de données associée de façon

opératoire avec le processeur,

dans lequel le dispositif de mémoire contient un corps d'instructions de programme incluant un modèle d'apprentissage automatique exécutable par le processeur et qui est configuré pour déterminer une probabilité d'interaction de l'utilisateur avec un élément de contenu, le modèle ayant été formé en utilisant un ensemble de vecteurs de caractéristiques de formation enrichies et des balises correspondantes d'événements d'interaction dérivées d'un ensemble de données appariées, généré à partir d'enregistrements concernant des événements de placement de contenu et des enregistrements concernant des événements d'interaction d'utilisateurs récupérés dans un entrepôt de données en ligne, dans lequel les événements de placement et d'interaction surviennent au cours d'un laps de temps défini

le corps des instructions de programme incluant par ailleurs des instructions qui, lorsqu'elles sont exécutées par le processeur, amènent l'appareil informatique à mettre en œuvre un procédé comprenant les étapes suivantes :

la réception, d'informations via l'interface de communications de données concernant l'encart pour le placement de contenu en ligne et des informations concernant un utilisateur auquel le contenu de l'encart de placement de contenu en ligne sera affiché ;

le calcul d'un vecteur de caractéristiques de probabilité enrichies, basé sur un élément de contenu sélectionné pour le placement à l'intérieur de l'encart de placement de contenu en ligne, les informations relatives à l'utilisateur et les informations relatives à l'encart de placement de contenu en ligne correspondant

l'exécution d'un modèle d'apprentissage automatique pour déterminer une probabilité d'interaction de l'utilisateur avec l'élément de contenu sélectionné sur la base du vecteur de caractéristiques de probabilité enrichies.

[0015] Dans des modes de réalisation de l'invention, le modèle d'apprentissage automatique est un modèle linéaire généralisé comprenant une pluralité de coefficients de modèle et en particulier le modèle d'apprentissage automatique peut être un modèle de régression logistique.

[0016] De façon avantageuse, la pluralité des coefficients de modèle peut être stockée dans une structure de donnée de dictionnaire dans laquelle chaque entrée est définie par une clé et une valeur de coefficient, dans laquelle chaque clé comprend une représentation hachée d'une concaténation d'un nom de caractéristique correspondant et d'une valeur de caractéristique correspondante. L'utilisation de ce type de structure de donnée permet le calcul d'une prédiction de la probabilité d'une interaction d'un utilisateur très rapidement, p. ex. en 30 ms. Les instructions de programme peuvent ainsi amener l'appareil informatique à mettre en œuvre l'étape d'exécution du modèle d'apprentissage automatique en :

- général, pour chaque valeur de caractéristique du vecteur de caractéristiques de probabilité enrichies, une clé correspondante ; et
- recupérant à partir de la structure de donnée de dictionnaire, pour chaque clé générée, une valeur de coefficient correspondante ; et
- calculant, en utilisant le vecteur de caractéristiques de probabilité enrichies et les valeurs de coefficient récupérées, l'estimation de probabilité d'une interaction de l'utilisateur avec l'élément de contenu sélectionné.

[0017] Dans des modes de réalisation dans lesquels l'encart de placement de contenu en ligne est un encart publicitaire, les informations relatives à l'encart publicitaire et les informations relatives à l'utilisateur auquel le contenu dans l'encart publicitaire sera affiché sont reçues avec un message de demande d'enchère transmis à partir d'un serveur de bourse publicitaire ; l'élément de contenu comprend au moins une enchère pour le placement à l'intérieur de l'encart publicitaire et le corps des instructions de programme inclut par ailleurs des instructions qui, lorsqu'elles sont exécutées par le processeur, amènent l'appareil informatique à implémenter le procédé comprenant les autres étapes suivantes :

la transmission, au serveur de bourse publicitaire, d'un message de réponse à une enchère en réponse au message de demande d'enchère ;

la réponse à la réception, en provenance du serveur de bourse publicitaire, d'une notification d'enchère réussie, l'actualisation d'un entrepôt de données en ligne avec des données d'événements de placement concernant le placement l'élément de contenu ; et

la réponse à la réception d'une notification d'une interaction d'un utilisateur avec l'élément de contenu, l'actualisation de l'entrepôt de données en ligne avec les données d'événements d'interaction relatives à l'interaction de l'utilisateur avec l'élément de contenu.

[0018] Dans un autre aspect, l'invention fournit un appareil informatique qui implémente la formation d'un modèle d'apprentissage automatique configuré pour estimer la probabilité de l'interaction d'un utilisateur avec des éléments de contenu, l'appareil informatique comprenant :

un processeur ;

au moins un dispositif de mémoire accessible par le processeur ; et  
un entrepôt de données accessible par le processeur,

dans lequel le dispositif de mémoire contient un corps d'instructions de programme incluant des instructions qui, lorsqu'elles sont exécutées par le processeur, amènent l'appareil informatique à mettre en œuvre un procédé comprenant les étapes suivantes :

l'accès à l'entrepôt de données pour récupérer des enregistrements concernant des événements de placement de contenu, et des enregistrements concernant des événements d'interaction d'un utilisateur, dans lequel les événements de placement et d'interaction surviennent au cours d'un laps de temps défini ;

l'appariement des enregistrements d'événements de placement de contenu avec des enregistrements d'événements d'interaction récupérés afin de générer un ensemble de données appariées qui comprend une pluralité d'enregistrements, chaque enregistrement de l'ensemble de données appariées incluant un ensemble de valeurs de caractéristiques brutes dérivées d'un événement de placement de

contenu avec une balise d'événement d'interaction indiquant si oui ou non un événement d'interaction correspondant à l'événement de placement de contenu est survenu ;

le calcul, à partir des valeurs de caractéristiques brutes, d'un ensemble correspondant de vecteurs de caractéristiques de formation enrichies ; et

la formation du modèle d'apprentissage automatique en utilisant les vecteurs de caractéristiques de formation et les balises d'événements d'interaction correspondantes.

[0019] Dans des modes de réalisation de l'invention, le modèle d'apprentissage automatique est un modèle de régression logistique et les instructions de programme amènent l'appareil informatique à implémenter l'étape de formation du modèle d'apprentissage automatique en utilisant la régression logistique régularisée avec l'algorithme d'apprentissage proximal 'Amorce régularisée de suivi' (Follow-The-Regularised-Leader) (FTRL).

[0020] Le corps des instructions de programme inclut par ailleurs des instructions qui, lorsque sont exécutées par le processeur, amènent l'appareil informatique à implémenter le procédé comprenant d'autres étapes, pour chaque coefficient de la pluralité des coefficients de modèles :

la génération d'une clé comprenant une représentation hachée d'un nom de caractéristique et d'une valeur de caractéristique correspondant au coefficient, et

le stockage, dans une structure de donnée de dictionnaire, d'une valeur de coefficient associée à la clé,

dans lequel la structure de donnée de dictionnaire comprend un chiffrage efficace du modèle d'apprentissage automatique.

[0021] Le corps des instructions de programme inclut des instructions qui, lorsqu'elles sont exécutées par le processeur, amènent l'appareil informatique à exécuter de façon récurrente les étapes suivantes : l'accès à l'entrepôt de données en ligne, l'appariement des enregistrements récupérés d'événements de

placement de contenu ainsi que les enregistrements récupérés d'événements d'interaction, le calcul des vecteurs de caractéristiques de formation enrichies et la formation (406) du modèle d'apprentissage automatique pour actualiser le modèle d'apprentissage automatique.

[0022] Dans un autre aspect, l'invention fournit un programme d'ordinateur comprenant un code de programme pour exécuter les étapes du procédé selon le premier aspect lorsque ledit programme est exécuté sur un ordinateur. Les instructions de code de programme peuvent par exemple être stockées sur des supports tangibles lisibles par ordinateur.

[0023] À partir de la description qui suit divers modes de réalisation, d'autres aspects, avantages et caractéristiques, des modes de réalisation de l'invention seront apparents aux hommes de métier spécialisés dans les domaines pertinents. On notera cependant, que l'invention n'est pas limitée aux modes de réalisation décrits qui sont fournis à titre d'illustration des principes de l'invention tels qu'ils sont définis dans les déclarations à venir et dans les revendications en annexe et pour assister les hommes de métier à mettre en pratique ces principes.

## BRÈVE DESCRIPTION DES DESSINS

[0024] Des modes de réalisation de l'invention seront maintenant décrits par rapport aux dessins accompagnants dans lesquels des numéros référentiels font référence à des caractéristiques similaires et dans lesquels :

La figure 1 est un diagramme schématique illustrant un système exemplaire en réseau pour mettre en œuvre l'invention ;

La Figure 2 montre une chronologie des communications entre un dispositif utilisateur, un serveur Web et un serveur de bourse publicitaire avec une DSP pour mettre en œuvre l'invention ;

La Figure 3 est un diagramme bloc illustrant de façon schématique un nombre de modules de code, comprenant un moteur de prédiction de l'interaction d'un utilisateur en ligne, pour mettre en œuvre l'invention ;

La Figure 4 montre un diagramme séquentiel d'un procédé d'actualisation d'un modèle d'apprentissage automatique en ligne pour mettre en œuvre l'invention ;

La Figure 5 montre un diagramme séquentiel d'un procédé d'ingénierie et d'optimisation d'hyperparamètres de modèle selon un mode de réalisation de l'invention ;

La Figure 6 illustre par un diagramme séquentiel un procédé pour le fonctionnement d'un module d'enchère en temps réel selon un mode de réalisation de l'invention ;

Les Figures 7(a) et 7(b) sont des diagrammes qui illustrent la performance d'un module d'enchère en temps réel pour mettre en œuvre l'invention.

## DESCRIPTION DÉTAILLÉE DES MODES DE RÉALISATION

[0025] La Figure 1 est un diagramme bloc qui illustre un système exemplaire en réseau 100 incluant du côté de la demande, un serveur 102 d'une plateforme (DSP) qui est configuré pour implémenter un procédé d'enchère pour le placement de contenu publicitaire selon un mode de réalisation de l'invention. Le serveur DSP 102 peut comprendre un système informatique ayant une architecture conventionnelle. En particulier, le serveur DSP 102, tel qu'il est illustré, comprend un processeur 104. Le processeur 104 est associé de façon fonctionnelle avec un dispositif de stockage/mémoire non volatile 106, p. ex. via un ou plusieurs bus de données/adresses 108 ainsi qu'illustré. Le stockage non volatile 106 peut être un disque dur et/ou peut inclure une mémoire non volatile à état défini, telle qu'une mémoire morte ROM, une mémoire flash, un disque dur électronique SSD ou autre stockage similaire. Le processeur 104 sert aussi d'interface avec le stockage volatile 110 tel que la mémoire à accès aléatoire RAM qui contient des instructions de programme et des données temporaires relatives au fonctionnement du serveur DSP 102.

[0026] Dans une configuration conventionnelle, le dispositif de stockage 106 conserve un programme connu et un contenu de données pertinentes avec le

fonctionnement normal du serveur DSP 102. Par exemple, le dispositif de stockage 106 peut contenir des programmes de système d'exploitation et des données ainsi que d'autres logiciels d'application exécutables, nécessaires pour les fonctions voulues du serveur d'authentification 102. Le dispositif de stockage 106 contient aussi des instructions de programme qui, lorsqu'elles sont exécutées par le processeur 104, amènent le serveur DSP 102 à effectuer des opérations relatives à un mode de réalisation de la présente invention, lesquelles sont décrites de façon plus détaillée ci-dessous en référence aux Figures 2 et 6 en particulier. Dans le fonctionnement, les instructions et les données stockées sur le dispositif de stockage 106 sont transférées à la mémoire volatile 110 pour une exécution à la demande.

[0027] Le processeur 104 est aussi associé fonctionnellement à une interface de communications 112 d'une manière conventionnelle. L'interface de communications 112 facilite l'accès à un réseau étendu de communications de données, tel que l'Internet 116.

[0028] Dans la pratique, le stockage volatile 110 contient un corps d'instructions de programme correspondant 114 transféré à partir du dispositif de stockage 106 et qui est configuré pour effectuer le traitement et d'autres opérations afin de mettre en œuvre les caractéristiques de la présente invention. Les instructions de programme 114 comprennent une contribution technique spécifique à l'art, conforme à l'invention décrite par la suite ci-dessous.

[0029] Concernant l'aperçu global du serveur DSP 102 et d'autres systèmes de traitement et dispositifs décrits dans cette spécification, les termes tels que 'processeurs', 'ordinateur' et ainsi de suite, sauf si le contexte exige autrement, doivent être interprétés comme faisant référence à une gamme d'implémentations possibles des dispositifs, appareils et systèmes comprenant une combinaison de logiciels et de matériel. Cela inclut des dispositifs à processeur unique et à processeurs multiples, et un appareil, incluant les dispositifs portables, les ordinateurs de bureau et divers types de systèmes de serveur incluant du matériel collaboratif et des plateformes logicielles qui peuvent

cohabiter ou être distribuées. Les processeurs physiques peuvent inclure des unités centrales de traitement (CPUs), à usage général, des processeurs de signaux numériques, des unités de traitement graphique (GPUs) et/ou d'autres dispositifs matériels appropriés pour une exécution efficace des programmes et des algorithmes nécessaires. Les systèmes informatiques peuvent inclure des architectures conventionnelles d'ordinateur personnel ou d'autres plateformes matérielles à usage général. Les logiciels peuvent inclure des logiciels de sources libres et/ou des logiciels de systèmes d'exploitation vendus sur le marché en combinaison avec diverses applications et programmes de services. Autrement, les plateformes de traitement ou de calcul peuvent comprendre du matériel et/ou des architectures de logiciel adaptés sur mesure. Pour une extensibilité améliorée, les systèmes de traitement et de calcul peuvent comprendre des plateformes d'informatique en nuage (cloud) permettant aux ressources matérielles physiques d'être affectées de façon dynamique en réponse aux demandes de services. Alors que toutes ces variations sont comprises dans la portée de la présente invention, pour faciliter l'explication et la compréhension, les modes de réalisation exemplaires décrits dans les présentes sont basés sur des plateformes informatiques à usage général et un processeur unique, des plateformes de système d'exploitation couramment disponibles et/ou des produits de consommateurs largement disponibles, tels que des ordinateurs de bureau (PCs), des ordinateurs compacts ou ordinateurs portables (PCs), des smartphones, des tablettes informatiques et ainsi de suite.

[0030] En particulier, le terme 'unité de traitement' est utilisé dans cette spécification (y compris les revendications) en référence à toute combinaison appropriée de matériel et de logiciels configurés pour effectuer une tâche particulière définie telle que l'accès et le traitement des données en ligne ou hors connexion, l'exécution des étapes de formation d'un modèle d'apprentissage automatique ou l'exécution des étapes de prédiction d'un modèle d'apprentissage automatique. Une telle unité de traitement peut comprendre un module de code exécutable qui s'exécute à un emplacement unique sur un dispositif de traitement unique ou qui peut comprendre des modules collaboratifs de code exécutable qui s'exécutent à de multiples emplacements et/ou sur de multiples dispositifs de

traitement. Par exemple, dans certains modes de réalisation de l'invention, le traitement d'une classification et d'une décision d'enchère peut être effectué entièrement par un code qui s'exécute sur le serveur DSP 102, alors que pour d'autres modes de réalisation le traitement correspondant peut être effectué d'une manière distribuée sur une pluralité de serveurs DSP.

[0031] Les composants logiciels, p. ex. les instructions de programme 114, pour mettre en œuvre les caractéristiques de l'invention peuvent être développés en utilisant tout langage de programmation approprié, environnement de développement ou combinaisons de langage et d'environnement de développement qui seront familiers aux hommes de métier spécialisés dans l'ingénierie logicielle. Par exemple, un logiciel approprié peut être développé en utilisant le langage de programmation C, le langage de programmation Java, le langage de programmation C++, le langage de programmation Go et/ou une gamme de langages adaptés à l'implémentation de services sur réseau, ou basés sur le Web, tels que JavaScript, HTML, PHP, ASP, JSP, Ruby, Python, Perl, et ainsi de suite. Ces exemples ne se veulent pas restrictifs et on notera que des langages ou des systèmes de développement pratiques peuvent être employés selon les exigences des systèmes. Les descriptions, diagrammes bloc, organigrammes et autres qui sont présentés dans cette spécification sont fournis à titre d'exemple pour permettre aux hommes de métier spécialisés dans l'ingénierie logicielle et l'apprentissage automatique de comprendre et d'apprécier les caractéristiques, la nature et la portée de l'invention, et pour mettre un ou plusieurs modes de réalisation de l'invention en pratique par l'implémentation de code logiciel adapté conformément à cette divulgation sans avoir à y apporter une ingéniosité inventive additionnelle.

[0032] Revenant à la Figure 1, le système 100 comprend par ailleurs des serveurs DSP additionnels, p. ex., 118, 120 qui, lors de l'utilisation, sont en concurrence avec le serveur DSP 102 pour enchérir sur le placement d'un contenu publicitaire à l'intérieur des encarts publicitaires offerts via un serveur de bourse publicitaire 122. Le serveur de bourse publicitaire 122 implémente une place de marché numérique qui permet aux annonceurs et aux éditeurs de sites

Web et d'autre contenu en ligne d'acheter et de vendre des espaces publicitaires sous forme d'une enchère en ligne en temps réel dans laquelle chaque serveur DSP 102, 118, 120 est un enchérisseur automatique ultrarapide. Le serveur de bourse publicitaire 122 comprend une base de données 124 dans laquelle il conserve les détails des fournisseurs de contenu en ligne (les serveurs Web) et des annonceurs (DSPs) dans le but de faire fonctionner une place de marché publicitaire numérique. Les fonctions des plateformes de bourse publicitaire telles que DoubleClick™ (propriété de Google™), AppNexus™, Microsoft™ Ad Exchange™ et OpenX™, sont bien connues et ne seront donc pas décrites de façon plus détaillée dans les présentes sauf si cela s'avère nécessaire pour illustrer de façon adéquate le fonctionnement des modes de réalisation de l'invention.

[0033] Le système 100 inclut par ailleurs des dispositifs de terminaux d'utilisateurs illustrés par le dispositif terminal 126. Les dispositifs de terminal 126 peuvent être par exemple des PCs de bureau ou portables, des smartphones, des tablettes ou d'autres dispositifs informatiques personnels et chacun comprend un processeur 128 qui est en interface, p. ex., via le bus d'adresses/données 130, avec le stockage volatile 132, le stockage non volatile 134 et au moins une interface de communications de données 136. Le processeur 128 est aussi en interface avec une ou plusieurs interfaces d'entrée/sortie (I/O) 140. Le stockage volatile 132 contient des instructions de programme et des données transitoires relatives au fonctionnement du dispositif d'un terminal 126.

[0034] Le stockage des dispositifs de terminaux 132, 134 peut contenir un programme et un contenu de données pertinents avec le fonctionnement normal du dispositif 126. Cela peut inclure des programmes de système d'exploitation et des données (p. ex. associés à Windows, Android, IOS, MacOS, Linux ou autre système d'exploitation) ainsi que d'autres logiciels d'application exécutables qui de façon générale n'ont pas de lien avec la présente invention. Le stockage 132 inclut aussi des instructions de programme 138 qui, lorsqu'elles sont exécutées par le processeur 128, permettent au dispositif de terminal de fournir un accès au

contenu en ligne à un utilisateur. Alors que beaucoup d'applications sont connues pour apporter cet accès, par simplicité dans la description présente il est supposé que les instructions de programme 138 mettent en œuvre un navigateur Web ayant une interface graphique d'utilisateur (GUI) présentée via l'interface I/O 140.

[0035] Par conséquent, dans l'hypothèse où un utilisateur du dispositif terminal 126 accéderait au serveur Web 142, un affichage 144 d'une page Web correspondante serait généré via le dispositif (UI) 140. L'affichage 144 inclut un contenu du site Web 146 et un ou plusieurs encarts publicitaires, p. ex. 148, 150. Comme illustré par ailleurs, chaque encart publicitaire 148, 150 peut comprendre une pluralité 'd'offres' spécifiques pour le compte d'un annonceur. Ces offres sont arrangées de manière courante sur un tracé de grille indiqué par les rectangles en ligne pointillée 148a, 148 b, 148c, 150a, 150 b, 150c dans la Figure 1. Un nombre d'étapes de communications peut ensuite prendre place afin de remplir ces encarts (c.-à-d. de fournir aux annonceurs en ligne des impressions de publicité à l'intérieur de l'affichage de la page Web 144. Ces étapes de communications vont maintenant être décrites en référence à la chronologie 200 illustrée dans la Figure 2.

[0036] Initialement, le terminal utilisateur 126, via l'application de navigateur Web en cours d'exécution 138 qui est réceptive à une entrée de l'utilisateur, transmet au 202 une requête HTTP au serveur Web 142 qui inclut une URL du contenu Web désiré. Le serveur Web 142 répond en transmettant au 204 le contenu, p. ex. une page Web en format HTML, au dispositif utilisateur 126. Les hommes de métier spécialisés dans l'art de la programmation web noteront que le remplissage complet et le rendu de l'affichage de pages Web 144 peut nécessiter de multiples requêtes et réponses et peut impliquer plusieurs autres transactions avec le serveur Web 142 et/ou avec d'autres serveurs en ligne, tels que les serveurs de réseau de distribution de contenu (CDN) et d'autres serveurs fournissant un contenu intégré. Par simplicité et pour faciliter l'attention sur les caractéristiques de mise en œuvre des communications de la présente invention, toutes ces transactions additionnelles connues sont représentées par une seule communication exemplaire 206 dans la Figure 2.

[0037] Afin d'obtenir du contenu publicitaire pour remplir les encarts 148 150, la page Web transmise par le serveur Web 142 au dispositif utilisateur 126 inclut typiquement une référence hypertexte ('href') qui instruit le navigateur 138 pour récupérer le contenu sur le serveur de la bourse publicitaire 122 en conformité avec une interface de programmation d'application (API) définie et fournie par l'exploitant concerné du serveur 122. Par conséquent, le dispositif utilisateur 126 transmet au 208 une requête HTTP au serveur de bourse publicitaire 122. La requête inclut des informations de site Web et des informations d'utilisateur relatives à l'utilisateur du dispositif terminal 126. Des informations d'utilisateur disponibles peuvent inclure des informations que le serveur Web 142 a rassemblées et peuvent inclure des informations du côté du client telles que l'identité du dispositif et du navigateur ainsi que des détails techniques, l'identification des informations, le contenu des cookies de navigateur et ainsi de suite. Plusieurs mécanismes en ligne pour la collecte, la conservation et le suivi des informations du dispositif et de l'utilisateur sont bien connus et disponibles aux hommes de métier spécialisés dans l'art de la programmation Web, par conséquent, ils ne seront pas décrits de façon détaillée dans la présente.

[0038] Le serveur de bourse publicitaire 122 reçoit la requête, identifie les serveurs DSP pertinents, 102, 118, 120 dans sa base de données 124 et transmet au 210 des messages indiquant une demande d'enchère pour chaque serveur DSP sélectionné. Un tel message de demande d'enchère, incluant des informations d'utilisateur et de site, est reçu par le serveur DSP 102 pour mettre en œuvre la présente invention qui exécute un processus 212 conformément à sa programmation spécifique 114 afin de prédire une probabilité d'interaction de l'utilisateur avec une publicité sélectionnée incluant une ou plusieurs offres, placées à l'intérieur d'un ou de plusieurs encarts disponibles 148, 150, et arriver à une décision d'enchère. Dans le cas où une décision est prise d'enchérir pour l'impression offerte et qu'une valeur d'enchère est déterminée, le serveur DSP 102 transmet ensuite au 214 l'enchère au serveur de bourse 122.

[0039] Le serveur de bourse 122 reçoit toutes les enchères transmises en provenance des serveurs DSP, incluant le serveur 102 et sélectionne une

enchère gagnante. Il récupère ensuite le contenu de l'annonce correspondant à l'enchère gagnante de sa base de données 124 et transmet au 216 le contenu de l'annonce au dispositif utilisateur 126 pour un rendu à l'intérieur de l'encart publicitaire correspondant, p. ex. 148 ou 150.

[0040] Il est bien connu que la vitesse de chargement d'une page est une caractéristique importante d'un site Web du point de vue de l'utilisateur et qu'il n'est pas désirable que le temps requis pour le chargement d'une page Web complète soit excessif. Il est typiquement préférable que le temps de chargement n'excède pas quelques secondes, p. ex. 3 secondes, au 218. Il y a, comme décrit ci-dessus, plusieurs étapes nécessaires pour servir complètement tout le contenu d'une page Web complexe qui peut impliquer de multiples serveurs dans l'ensemble de l'Internet mondial. Il est donc primordial que la durée du processus d'enchère facilité par le serveur de bourse 202 soit strictement limitée. Il est actuellement jugé désirable que le serveur DSP 102 prenne une décision d'enchère en l'espace de pas plus de quelques dixièmes de millisecondes, par exemple en moins de 30 millisecondes, au 220. Cette décision doit être prise avec des informations d'utilisateur limitées et compte tenu du fait qu'une mauvaise décision peut avoir des conséquences substantielles pour l'annonceur. Par exemple, si le serveur DSP détermine à tort que l'utilisateur est une cible désirable pour une annonce particulière (c.-à-d. calcule un 'faux positif') il peut placer une enchère gagnante substantiellement élevée et subir un coût réel avec peu ou aucune chance de retour. Inversement, si le serveur DSP détermine que l'utilisateur n'est pas une cible désirable pour l'annonce (c.-à-d. calcule un 'faux négatif'), il peut ne pas placer d'enchère ou placer une enchère basse perdante et faire manquer une occasion à l'annonceur d'obtenir une impression avec une véritable chance de retour.

[0041] Afin d'atteindre une prise de décision qualitative à grande vitesse dans le contexte des services de réservation de voyage, des modes de réalisation de la présente invention utilisent une approche d'apprentissage automatique. Pour faciliter par ailleurs la compréhension de cette approche, la Figure 1 est maintenant référencée, dans laquelle le système 100 inclut par ailleurs un

serveur d'apprentissage automatique (Serveur ML) 152 qui est configuré pour traiter des données brutes relatives au placement du contenu (c.-à-d. les clics de l'utilisateur sur des annonces/offres), générer des ensembles de données de formation pour un modèle d'apprentissage automatique et former le modèle d'apprentissage automatique pour un déploiement sur le serveur DSP 102. Les étapes de traitement, de formation et de déploiement sont décrites de façon plus détaillée ci-dessous en référence aux Figures 3 et 4 et peuvent être mises en œuvre de façon continue, périodiquement et/ou à la demande afin de maintenir l'actualité du modèle d'apprentissage automatique.

[0042] Comme avec le serveur DSP 102, le serveur ML 152 peut comprendre un système informatique ayant une architecture conventionnelle, p. ex. comprenant un processeur 154 qui est associé de façon fonctionnelle avec un dispositif de mémoire/stockage non volatile 156, via un ou plusieurs bus de données/adresses 158, tel qu'illustré. Le processeur 154 sert aussi d'interface au stockage volatile 160 qui contient les instructions de programme et les données transitoires concernant le fonctionnement du serveur ML 152. Conventionnellement, le dispositif de stockage 156 contient des programmes de systèmes d'exploitation et des données ainsi que d'autres logiciels d'application exécutables nécessaires aux fonctions voulues du serveur ML 152 et inclut des instructions de programmes qui, lorsqu'elles sont exécutées par le processeur 154, amènent le serveur ML 152 à mettre en œuvre les opérations concernant un mode de réalisation de la présente invention telles qu'elles seront décrites de façon plus détaillée ci-dessous en référence aux Figures 3 et 4 en particulier. En pratique, les instructions et les données stockées sur le dispositif de stockage 156 sont transférées à la mémoire volatile 150 pour une exécution à la demande. En outre, le processeur 154 est associé fonctionnellement à une interface de communications 162 d'une manière conventionnelle, fournissant un accès à l'Internet 116.

[0043] En pratique, le stockage volatile 160 contient un corps d'instructions de programme 164 correspondant, transféré à partir du dispositif de stockage 156 et configuré pour effectuer les étapes de traitement, de formation et de déploiement

pour mettre en œuvre la présente invention. Les instructions de programme 164 comprennent une autre contribution technique spécifique à l'art conformément à l'invention.

[0044] Le système 100 inclut par ailleurs au moins une base de données 166 qui est configurée pour stocker des données historiques brutes concernant le placement du contenu (c.-à-d. annonces/offres) avec les interactions d'un utilisateur (l'utilisateur clique sur des annonces/offres). Le volume de ces données peut être très large dans les laps de temps présentant un intérêt, par exemple un mois ou plus. Par exemple, dans un déploiement particulier en temps réel il a été découvert qu'un registre de données pour une seule journée comprenait de l'ordre de 20 millions de lignes (c.-à-d. d'événements de placement et d'interaction) ayant au total une taille de stockage de l'ordre de 10 Gb. Par conséquent, la base de données 166 est de préférence implémentée en utilisant des technologies qui sont optimisées pour un stockage, une récupération et une actualisation efficaces de très larges volumes de données (parfois appelées « mégadonnées ») pour de multiples serveurs de base de données et de dispositifs de stockage. Bien qu'un nombre de technologies libres et commerciales adaptées existe pour l'implémentation de la base de données 166, un mode de réalisation exemplaire expérimental a été mis en œuvre en utilisant l'architecture Apache Hadoop avec des données stockées en format Parquet sur HDFS (un système de fichier Hadoop distribué) et en utilisant Impala, un moteur d'interrogation similaire à SQL pour fournir une vitesse ultrarapide. L'implémentation a été testée et s'est avérée fournir une performance plus qu'adéquate pour le déploiement pratique en ligne des modes de réalisation de l'invention.

[0045] La base de données 166 est accessible à la fois au serveur DSP 102 et au serveur ML 152. Dans la Figure 1, l'accès logique est illustré par des flèches correspondantes. Dans un mode de réalisation pratique, l'accès physique entre la base de données 166 et les serveurs DSP et ML 102, 152 peut se faire via l'Internet 116 et/ou via d'autres liens ou réseaux de communications dédiés, tels que le réseau d'espace de stockage local (SAN). Le serveur DSP 102 est

configuré pour mettre à jour la base de données 166 en temps réel avec des données brutes relatives aux événements de placement et d'interaction. Le serveur ML 152 est configuré pour récupérer les données brutes de la base de données 166 et pour mettre en œuvre les étapes de traitement, de formation et de déploiement, sur la base des données récupérées, selon un mode de réalisation de l'invention.

[0046] Revenant à la Figure 2, d'autres opérations relatives à la mise à jour de la base de données 166 par le serveur DSP 102 sont illustrées. En particulier, dans l'éventualité où le serveur DSP 102 placerait une enchère gagnante et qu'un contenu d'annonce correspondant serait transmis au 216 au dispositif utilisateur 126, le serveur DSP 102 actualiserait au 222 la base de données 166, en ajoutant des données relatives au placement de l'annonce (c.-à-d. l'impression d'une annonce/offre). Le code associé à l'annonce est configuré de sorte que, dans l'éventualité où l'utilisateur interagirait par la suite avec l'annonce (c.-à-d. en cliquant dessus), le serveur DSP 102 recevrait directement ou indirectement une notification 224 de cet événement d'interaction. Ensuite, au 226, le serveur DSP 102 met à jour la base de données 166 avec des détails de l'événement d'interaction. De cette façon, la base de données 166 est continuellement actualisée avec des données brutes relatives à tous les événements de placement et d'interaction connus par le serveur DSP 102.

[0047] La Figure 3 est un diagramme bloc illustrant de façon schématique un nombre de modules de code qui ensemble comprennent un moteur de prédiction d'interaction (300) d'utilisateur en ligne pour mettre en œuvre l'invention. L'implémentation du moteur de prédiction d'interaction de l'utilisateur 300 est distribuée entre le serveur ML 152 et le serveur DSP 102 comme le montrent les cases en pointillés dans la Figure 3. Trois modules de code constituent les composants de serveur ML du moteur 300, notamment un module d'appariement 302, un module d'enrichissement de caractéristiques 304 et un module d'apprentissage automatique 306. Ces trois modules sont tous implémentés par les instructions de programmes 164 qui s'exécutent sur le

serveur ML 152. La fonctionnalité implémentée dans chacun de ces modules sera maintenant décrite de façon plus détaillée.

[0048] Le but du module d'appariement 302 est d'apparier les événements de placement (c.-à-d. l'affichage des annonces et des offres à l'intérieur des annonces dans les encarts publicitaires 148, 150 de l'affichage 144 du dispositif utilisateur 126) avec des événements d'interaction subséquents (c.-à-d. les occurrences de clics d'un utilisateur sur une offre à l'intérieur d'une annonce placée sur l'affichage 144 du dispositif utilisateur 126). L'appariement permet le placement des événements devant être balisés pour avoir 'reçu un clic' ou 'ne pas avoir reçu de clic', de sorte qu'ils peuvent être utilisés par le module d'apprentissage automatique 306 pour la formation d'un modèle d'apprentissage automatique contrôlé afin de prédire les événements d'interaction de l'utilisateur sur la base du placement des données d'événements. En outre, l'appariement permet le placement de données d'événements devant être combinées avec des données d'événements d'interaction correspondants pour créer un enregistrement des annonces ayant fait l'objet de clics contenant toutes les informations disponibles concernant le placement et l'interaction.

[0049] L'appariement présente un défi parce qu'il n'existe pas de lien explicite entre un événement de placement (impression d'annonce) et une interaction d'utilisateur par la suite (un clic sur l'annonce). Comme illustré par la chronologie 200 de la Figure 2, une interaction d'un utilisateur peut survenir à tout moment après le placement, p. ex. après un délai substantiel. Puisque le taux d'occurrence des événements d'interaction et/ou de placement peut être très élevé (p. ex. des centaines de milliers de fois par seconde), dans un système opérationnel, l'appariement des événements d'interaction et de placement peut devenir largement séparé dans la base de données 166. En outre, le taux d'événements d'interaction peut être très bas, p. ex. il est généralement rapporté que le taux de clics (CTR) pour les publicités affichées sur le Web est de l'ordre de 0,05 %. Par ailleurs, il est souhaitable de lier les événements d'interaction et de placement au niveau de l'offre, plutôt qu'uniquement au niveau de l'annonce.

[0050] L'approche générale employée pour l'appariement dans les modes de réalisation de l'invention est d'identifier dans la base de données 166 des événements de placement et des événements subséquents d'interaction dans une fenêtre de temps prédéterminée qui ont un ensemble sélectionné de paramètres correspondants. La fenêtre de temps doit avoir une durée suffisante pour capturer une majorité substantielle des interactions dans leur ensemble et le nombre et le choix des paramètres doivent être suffisants pour assurer un appariement unique dans une majorité substantielle de cas. Un appariement parfait peut être difficile à atteindre parce qu'il est impossible de savoir si, ou quand, une interaction va survenir. Une fenêtre de temps d'une durée plus longue capturera des interactions qui surviennent après des délais plus longs, mais augmentera aussi le risque d'appariements erronés lorsque, par exemple, un utilisateur interagit avec une annonce présentée par la suite ayant des paramètres similaires. De façon similaire le risque d'appariement erroné peut être réduit en utilisant un ensemble plus large de paramètres sélectionnés pour faire la distinction entre des annonces présentées, au risque de rendre le processus d'appariement plus complexe.

[0051] Dans un mode de réalisation expérimental, l'invention a été implémentée dans le contexte d'un serveur DSP d'un domaine spécifique fonctionnant pour le compte des annonceurs en utilisant des données d'événements capturées à partir d'un système opérationnel. Une approche heuristique a été adoptée pour concevoir le module d'appariement avec un nombre d'expériences conduites pour déterminer une fenêtre de temps appropriée et un ensemble sélectionné de paramètres. Une fenêtre de temps de 80 secondes s'est avérée efficace en combinaison avec les paramètres d'événements correspondants suivant :

- un identifiant d'utilisateur unique (suivi via un cookie de navigateur) ;
- un identifiant d'annonceur ;
- un identifiant d'éditeur (c.-à-d. le réseau de distribution/bourse publicitaire via lequel l'annonce a été placée)

- le format de l'offre sélectionnée par un clic (p. ex. la largeur et la hauteur graphique de l'offre en pixels) ;
- le type de produit d'annonce ;
- la réserve de produits d'annonce ;
- le segment d'utilisateur (une combinaison du segment de produit d'un utilisateur, basée sur un produit tel qu'un vol, un hôtel ou un restaurant consulté précédemment par l'utilisateur et un segment de temps d'utilisateur indiquant la durée écoulée depuis la dernière activité de l'utilisateur) ;
- l'URL du site
- la visibilité de l'encart publicitaire
- le dispositif utilisateur ;
- une mesure de la distance entre une destination (lieu) au sujet duquel l'utilisateur recherchait des informations et une destination qui a fait l'objet d'une offre spécifique ; et
- une clé d'encart publicitaire (un identifiant stable pour la combinaison : éditeur, encart publicitaire et page).

[0052] Dans le mode de réalisation exemplaire, l'appariement est effectué en utilisant une interrogation Impala SQL pour sélectionner et joindre des tableaux d'enregistrement d'événements de placement et d'interaction sur les valeurs de champs correspondants aux paramètres listés ci-dessus. Spécifiquement, les enregistrements de placement sont joints à la gauche des enregistrements d'interaction, de sorte que le tableau qui en résulte inclut un rang pour chaque événement de placement. Chaque rang comprend un ensemble de valeurs de caractéristiques brutes dérivées des événements appariés, ainsi qu'un indicateur indiquant si oui ou non un événement d'interaction, c.-à-d. un clic sur une annonce/offre, est survenu. Le tableau de données appariées est entré dans le module d'enrichissement de caractéristiques 304.

[0053] La fonction du module d'enrichissement de caractéristiques 304 est de dériver, à partir des valeurs de caractéristiques brutes dans le tableau de données appariées, générées par le module d'appariement 302, un ensemble correspondant de vecteurs de caractéristiques enrichies pour une utilisation par le module d'apprentissage automatique 306. Un processus pour déterminer un ensemble approprié de caractéristiques enrichies ((c.-à-d. une ingénierie de caractéristiques) est décrit de façon plus détaillée ci-dessous en référence à la Figure 5. Dans la Figure 3, les définitions de caractéristiques enrichies pour une utilisation par le module d'enrichissement de caractéristiques 304 sont représentées comme étant stockées dans un fichier 310 à l'intérieur de l'entrepôt de données 308, cependant ceci doit être considéré comme une commodité schématique. Dans un mode de réalisation pratique, les définitions de caractéristiques peuvent être stockées de cette façon, elles peuvent être compilées dans un module de code et liées au module d'enrichissement de caractéristiques 304 ou elles peuvent faire l'objet d'un codage dur (hard-coded) dans le module d'enrichissement de caractéristiques. On notera que chacune de ces options d'implémentation (et d'autres qui seront apparentes aux hommes de métier) offre potentiellement un compromis différent entre : la flexibilité, la complexité de code et la vitesse d'exécution.

[0054] Dans le mode de réalisation exemplaire, toutes les caractéristiques enrichies sont de type catégorique (c.-à-d. elles prennent une valeur parmi un nombre de valeurs discrètes) et sont encodées à chaud (one hot) Les vecteurs de caractéristiques résultants sont donc généralement relativement épars et comprennent des éléments binaires. Par ailleurs, un vecteur de caractéristiques correspond à une offre à l'intérieur d'une publicité présentée à un utilisateur et il est associé à une balise binaire indiquant si oui ou non l'utilisateur a interagi avec l'offre (c.-à-d. cliqué dessus). Le tableau résultant de vecteurs de caractéristiques et de balises est entré dans le module d'apprentissage automatique 306.

[0055] Le module d'apprentissage automatique 306 comprend un code de programme s'exécutant sur le serveur ML 152 qui est configuré dans le mode expérimental de réalisation exemplaire pour implémenter un modèle linéaire

généralisé. Spécifiquement, le module d'apprentissage automatique 306 du mode expérimental de réalisation exemplaire implémente un algorithme de régression logistique régularisée avec une fonction d'apprentissage proximal 'd'Amorce régularisée de suivi' (Follow-The-Regularised-Leader) (FTRL). De façon avantageuse cet algorithme d'apprentissage automatique est connu pour être efficace dans les cas impliquant des ensembles de données extrêmement déséquilibrés (sachant que seul environ 0,05 % des échantillons dans le tableau de vecteur de caractéristiques sont balisés comme ayant reçu un clic). Plus de détails sur l'algorithme et son application pour la prédiction de clics peuvent être trouvés dans l'ouvrage de H. Brendan McMahan *et al*, intitulé 'Ad Click Prediction : a View from the Trenches', *KDD' 13*, August 11–14, 2013, Chicago, Illinois, USA. L'algorithme a un nombre d'hyperparamètres qui peuvent être ajustés afin d'optimiser sa précision d'apprentissage sur les données de formation pour un problème spécifique. Un processus pour déterminer un ensemble approprié de valeur pour les hyperparamètres est décrit de façon plus détaillée ci-dessous en référence à la Figure 5. Dans la Figure 3, des valeurs fixes d'hyperparamètres pour une utilisation par le module d'apprentissage automatique 306 sont illustrées comme étant stockées dans un fichier 312 dans l'entrepôt de données 308. Cependant, comme on le remarquera, d'autres implémentations sont possibles telles que l'encodage fixe des paramètres dans le module d'apprentissage automatique 306.

[0056] L'exécution du module d'apprentissage automatique 306 sur les résultats d'un ensemble particulier de données donne lieu à la génération d'un modèle qui peut être exécuté par le serveur DSP 102 tel qu'il sera décrit de façon plus détaillée ci-dessous en référence à la Figure 6. En particulier, un modèle de régression logistique est totalement caractérisé par un ensemble de coefficients associés aux éléments d'entrée du vecteur de caractéristiques. Dans le mode de réalisation exemplaire, une représentation particulièrement efficace du modèle est employée permettant au serveur DSP 102 de calculer une prédiction de la probabilité d'une interaction d'utilisateur très rapidement, c.-à-d. dans la fenêtre de temps ciblé 220 de l'ordre de 30 ms, pour générer une décision d'enchère. Spécifiquement, les coefficients sont stockés dans une structure de donnée de

dictionnaire dans laquelle chaque entrée est définie par une clé et une valeur. La clé est une représentation hachée d'une concaténation de noms de caractéristiques (c.-à-d. l'étiquetage de colonnes dans le tableau de caractéristiques) et une valeur de caractéristiques correspondantes (c.-à-d. des valeurs catégoriques avant l'encodage à chaud [one-hot]). La valeur associée dans le dictionnaire est simplement le coefficient de modèle correspondant. Ce type de structure de donnée est connu pour fournir une consultation extrêmement rapide, particulièrement dans des ensembles de caractéristiques épars. En particulier, en utilisant des valeurs hachées, une limite sur le nombre de caractéristiques hachées peut-être imposée (un mécanisme parfois désigné sous le nom de 'lecture spéciale de hachage' (hashing trick)). Ce mécanisme peut être utilisé pour accélérer substantiellement la vitesse de consultation et le calcul, au risque de collisions possibles dans les valeurs de clé de dictionnaire. Cependant, de façon avantageuse, les effets statistiques de ces collisions peuvent être négligés en ce qui concerne la performance globale de l'algorithme.

[0057] Pour un déploiement au serveur DSP 102, le modèle de structure de donnée est sérialisé en format binaire (dans le mode de réalisation exemplaire, le format Python 'pickle' est utilisé) et stocké dans un fichier de modèle 314 dans l'entrepôt de données 308.

[0058] En pratique, le serveur ML 152 exécute les modules 302, 304, 306 de façon récurrente, p. ex. de façon continue, périodique ou à la demande. Cela est illustré par l'organigramme 400 montré dans la Figure 4. Des données brutes sont récupérées dans la base de données 166 à l'étape 402. Des modes de réalisation exemplaires utilisent une période prédéterminée de données récentes qui est considérée comme étant représentative du comportement des utilisateurs actuels en ligne du système 100. Par exemple, des données brutes issues de la période d'un mois la plus récente peuvent être employées. À l'étape 404, le module d'appariement 302 effectue l'appariement des événements de placement et d'interaction de la façon qui a été décrite. En pratique, les étapes de récupération 402 et d'appariement 404 peuvent être combinées en une seule interrogation, p. ex. une interrogation SQL Impala.

[0059] À l'étape 406, le serveur ML152 exécute le module d'enrichissement de caractéristiques qui utilisent les définitions de caractéristiques enrichies 310 pour calculer les vecteurs de caractéristiques enrichies correspondant aux données appariées. Celles-ci sont transférées au module d'apprentissage automatique 306 qui forme le modèle en utilisant les vecteurs de caractéristiques balisés et les hyperparamètres définis dans le fichier de configuration 312. Les coefficients du modèle résultant sont hachés, sérialisés et publiés au 410 dans le fichier de modèle 314.

[0060] Ensuite, optionnellement, le serveur ML attend au 412, avant de recommencer le processus à l'étape 402. Une sortie de la condition d'attente 412 peut être déclenchée par un nombre d'événements différents. Par exemple, le serveur ML peut être configuré pour faire fonctionner les modules 302, 304, 306 de façon périodique, p. ex. une fois par jour. Autrement, ou de surcroît, il peut être configuré pour faire fonctionner les modules 302, 304, 306 à la demande, p. ex. à la réception d'un signal provenant d'un contrôleur (non illustré) à l'intérieur du système 100. Dans certains modes de réalisation, le serveur ML peut faire fonctionner les modules 302, 304, 306 de façon continue actualisant ainsi le fichier de modèle 314 aussi souvent que possible selon le temps requis pour appairer les données, enrichir les caractéristiques et former le modèle. Dans une configuration exemplaire expérimentale, il s'est avéré que des mises à jour basées sur des lots de données de 30 minutes fournissait un compromis approprié entre la qualité du produit du module d'appariement 302 (c.-à-d. le besoin de faire un rapprochement entre les événements d'interaction et de placement de façon précise pour un bon ensemble de données de formation) et la réactivité au changement en temps réel dans le réseau de la bourse publicitaire (p. ex. les lancements de nouvelles campagnes, les entrées/sorties des concurrents, les changements dans la demande des utilisateurs pour des contenus et ainsi de suite).

[0061] Faisant maintenant référence à la Figure 5, où l'on voit un organigramme 500 représentant un processus d'ingénierie de caractéristiques et d'optimisation d'hyperparamètres de modèle selon un mode de réalisation de

l'invention. En pratique le processus 500 est partiellement automatisé et son fonctionnement est supervisé par des humains. Le développement de caractéristiques appropriées avec une forte capacité de prédiction et la sélection d'une gamme d'hyperparamètres de modèles adaptés demandent une expérience substantielle, du jugement, de la créativité et de l'ingéniosité, et dans la plupart des cas ne peuvent pas être entièrement automatisés de façon efficace.

[0062] Le processus 500 requiert un ensemble de données de test qui sont récupérées à l'étape 502 et qui peuvent être obtenues de la manière décrite ci-dessus par rapport à la fonctionnalité du module d'appariement 302. En particulier, des données peuvent être extraites de la base de données 166 pour une période de test sélectionnée en utilisant une interrogation SQL Impala de la même forme que celle qui est utilisée par le module d'appariement 302.

[0063] À l'étape 504, un ensemble de caractéristiques enrichies est défini et configuré. Dans le mode de réalisation exemplaire, cette étape demande l'application de jugement, de créativité et d'ingéniosité de la part d'un expert scientifique en données. En pratique, un nombre d'expériences ont été réalisées conformément au processus 500 et ont été renforcées par d'autres analyses de l'ensemble de données de test afin d'identifier un ensemble efficace de caractéristiques enrichies. À l'étape 506, les valeurs des caractéristiques enrichies définies sont calculées à partir de l'ensemble de données brutes de test.

[0064] À l'étape 508, un ensemble de valeurs d'hyperparamètres est sélectionné et un modèle d'apprentissage automatique est configuré avec les valeurs sélectionnées. À l'étape 510, le modèle résultant est formé en utilisant les données de test enrichies. Typiquement, une portion des données de test est retenue dans l'étape de formation 510 laquelle est ensuite utilisée dans une validation croisée à l'étape 512 pour évaluer la performance du modèle formé avec des données qui n'ont pas été vues pendant l'étape de formation 510.

[0065] La performance du modèle formé est ensuite évaluée à l'étape de décision 514 pour déterminer si oui ou non elle est acceptable, par exemple en atteignant un niveau de performance optimal ou suffisant. Le choix des critères pour évaluer la performance peut être important pour identifier un modèle acceptable. Des critères divers connus peuvent être utilisés, tels que la surface sous la courbe de fonctionnement du récepteur (AUROC), la perte de registre, ou Gini (qui s'apparente à AUROC). Dans le mode de réalisation exemplaire, une combinaison de Gini (qui prend des valeurs entre -1 et 1 et dont il est souhaité qu'elle soit aussi élevée que possible) a été utilisée pour évaluer la performance de modèles différents. Cette approche a été utilisée non seulement pour des hyperparamètres différents du modèle proximal FTRL sélectionné, mais aussi pour un nombre d'autres modèles incluant des arbres de décisions (une forêt aléatoire distribuée, des arbres gradients renforcés) l'algorithme Bayes naïf et des réseaux d'apprentissage profond qui ont été rejetés par la suite pour avoir fourni une performance inférieure sur les ensembles de données analysées.

[0066] Dans l'éventualité où la performance serait jugée inacceptable ou qu'un processus d'optimisation s'avère incomplet, à l'étape de décision 514 une autre décision 516 est prise quant à savoir si les hyperparamètres de modèle sont actualisés. La boucle de configuration des hyperparamètres qui en résulte, la formation et le test du modèle sont typiquement automatisés en utilisant un algorithme comme un quadrillage de recherche, ou une action similaire. Le rôle de l'expert scientifique en données qui supervise dans ce cas est de déterminer l'étendue appropriée pour le quadrillage des hyperparamètres.

[0067] Au cas où il n'y aurait plus d'exigences en matière de variation des hyperparamètres, une boucle extérieure, implémentée via la décision 518, permet de tester d'autres ensembles de caractéristiques enrichies. Si des sélections et des valeurs de modèles d'algorithme, d'hyperparamètres et de caractéristiques enrichies ont été épuisés sans identification d'un modèle acceptable alors on peut considérer que le processus 500 a échoué et il peut être nécessaire de revoir la stratégie. Cependant, pour les besoins du mode de réalisation exemplaire, le processus 500 a conduit à un modèle dont la performance est acceptable. Par

conséquent à l'étape 520, les définitions de caractéristiques enrichies et de modèles d'hyperparamètres identifiés sont écrites dans des fichiers de données 310, 312 dans l'entrepôt de données 308. Un résumé des caractéristiques enrichies développées via le processus 500 est présenté dans le tableau 1.

[0068] Faisant référence à nouveau à la Figure 3, le moteur de prédiction d'interaction d'un utilisateur en ligne 300 inclut un module d'enchère en temps réel 316 qui est implémenté dans les instructions de programme 114 s'exécutant sur le serveur DSP 102. Le module d'enchère en temps réel 316 utilise les définitions de caractéristiques enrichies 310 et la représentation du modèle formé 314. En particulier, le fonctionnement du module d'enchère en temps réel 316 est représenté par l'organigramme 600 illustré dans la Figure 6, dans lequel, à l'étape 602, les informations de sites et d'utilisateurs sont reçues (c.-à-d. via la transmission de 210 en provenance du serveur de bourse publicitaire 122. Cette information est utilisée à l'étape 604 pour calculer un vecteur de caractéristiques enrichies correspondant conforme aux définitions 310.

[0069] À l'étape 606, le module d'enchère en temps réel accède à la représentation du modèle qui, comme cela a été décrit, comprend un ensemble de coefficients stockés dans une structure de dictionnaire ultra-efficace permettant une consultation de coefficient rapide. Comme décrit ci-dessus, en référence à la Figure 4 en particulier, le modèle peut être actualisé de temps en temps par le serveur ML 152. La représentation du modèle 314 peut être stockée dans un support de stockage partagé 308 et être lue de façon asynchrone par le serveur DSP 102. Dans certains modes de réalisation, le serveur DSP peut conserver une copie de la représentation de modèle 314 en antémémoire pour un accès rapide, laquelle est actualisée lors de la mise à jour du fichier stocké par le serveur ML 152.

[0070] Le produit du modèle est une estimation de la probabilité d'une l'interaction d'un utilisateur avec une offre dans une publicité sélectionnée sur la base du vecteur de caractéristiques enrichies. Dans le mode de réalisation exemplaire, le produit est une valeur représentant une probabilité que l'utilisateur

va cliquer sur une offre dans l'annonce sélectionnée. Cette valeur est utilisée dans le processus de décision d'enchère à l'étape 608. Le processus 608 peut inclure de déterminer si oui ou non il faut enchérir et/ou peut déterminer un prix spécifique pour l'enchère de l'encart publicitaire disponible. Par exemple, un seuil peut être appliqué de sorte que si la valeur est en dessous du seuil aucune enchère ne peut être placée. Dans certains modes de réalisation, le montant d'une enchère peut être déterminé sur la base de la magnitude de la valeur, de sorte qu'un prix plus élevé est enchéri si le modèle indique une probabilité plus élevée que l'utilisateur va cliquer sur une offre dans l'annonce sélectionnée. Dans l'éventualité où une décision serait faite d'enchérir pour l'encart, le contrôle est orienté au 610 vers l'étape 612 dans laquelle les informations de l'enchère sont renvoyées au 214 aux serveurs de bourse publicitaire 122. Dans l'éventualité où l'enchère serait réussie, le contrôle serait orienté au 614 vers l'étape 616, dans laquelle la base de données 166 est mise à jour avec les détails de l'événement de placement.

[0071] Afin d'évaluer la performance du module d'enchère en temps réel 316 pour implémenter l'invention, un module expérimental a fonctionné en parallèle avec un nombre de modules implémentant un algorithme d'enchère conventionnel. Les résultats sont montrés dans les diagrammes des Figures 7(a) et 7(b)

[0072] En particulier, la Figure 7(a) est un diagramme 700 ayant un taux de clics (CTR) sur l'axe vertical 702 avec la performance correspondante de dix modules d'enchère représentés par une série de barres. Les barres 704 représentent la performance de neuf modules conventionnels d'enchère alors que la barre 706 représente la performance de l'enchérisseur expérimental représentant l'invention. Comme on peut le constater, l'enchérisseur expérimental a atteint un CTR d'environ cinq fois plus que la performance moyenne des enchérisseurs conventionnels.

[0073] La Figure 7(b) est un diagramme 708 ayant une marge, définie par le bénéfice net divisé par le coût, sur l'axe vertical 710. Les barres 712 représentent

la performance de neuf modules d'enchère conventionnels, lesquels ont tous fonctionné à perte. Cependant, l'enchérisseur expérimental, représenté par la barre 714, a été capable de fonctionner de façon profitable.

[0074] Les résultats des Figures 7(a) et 7(b) démontrent donc clairement la supériorité pratique et technique atteinte par l'invention par rapport aux procédés conventionnels de prédiction d'interaction d'un utilisateur en ligne.

[0075] On notera que bien que des modes de réalisation particuliers et des variations de l'invention ont été décrits dans les présentes, d'autres modifications et alternatives seront apparentes aux hommes de métier spécialisés dans les arts pertinents. En particulier, les exemples sont offerts à titre d'illustration des principes de l'invention et pour apporter un nombre de procédés et d'arrangements spécifiques pour mettre en œuvre ces principes. De façon générale, les modes de réalisation de l'invention s'appuient sur l'apport d'arrangements techniques dans lesquels la prise de décision automatisée en ligne en temps réel peut être effectuée sur la base des prédictions dérivées d'interaction d'utilisateur à partir d'un modèle d'apprentissage automatique formé qui utilise des données dérivées d'une base de données d'événements de placement et d'interaction. Les étapes techniques implémentées par les modes de réalisation exemplaires incluent l'appariement des événements pour générer des enregistrements combinés de placement/interaction qui sont balisés pour usage par des algorithmes d'apprentissage contrôlés, le calcul de vecteurs de caractéristiques enrichies pour l'apprentissage en ligne et la formation d'un modèle d'apprentissage automatique basé sur une actualisation continue de données d'événements afin de maintenir une représentation de modèle à jour et actualisée périodiquement ayant un format efficace utilisable par un module

d'enchère en temps réel pour prendre des décisions rapides, p. ex. en moins de 30 ms

TABLEAU 1 Résumé des caractéristiques enrichies

Nom de la caractéristique	Description de la caractéristique
ts_day_of_week	Le jour de la semaine (dim-sam) de l'événement de placement.
ts_hour_of_day	L'heure de la journée (00-23) de l'événement de placement.
ts_is_weekend	Si le placement de l'événement est survenu au cours d'un weekend.
ts_is_bank_holiday	Si l'événement de placement est survenu pendant un jour férié pour les banques dans le pays à partir duquel l'utilisateur a accédé au site.
publisher_id	L'identifiant de l'éditeur (c.-à-d. l'exploitant du serveur de bourse publicitaire).
advertiser_id	L'identifiant de l'annonceur.
offer_key	Un identifiant unique d'offre, créé en combinant id_annonceur (voir ci-dessus) avec d'autres champs de l'annonceur (type de produit et réserve de produit).
ad_dst_top199	Une destination associée à une offre. Limitée aux 199 premières destinations qui lors des expériences d'ingénierie de caractéristiques ont révélé la capture de 92 % de tous les clics.
fmt	Format d'une offre (largeur et hauteur de l'image d'offre dans l'encart publicitaire).
nb_offers_per_ad	Nombre d'offres incluses avec l'encart publicitaire.
mq_dst	Proximité/distance de la destination d'intérêt pour l'utilisateur et de la destination associée à une offre. Une valeur catégorique indiquant la proximité d'un appariement sur une échelle définie.
user_pseg	L'identifiant d'un segment de produit consulté précédemment par un utilisateur (p. ex. un vol, un hébergement, un restaurant).
user_tseg	L'identifiant d'un segment de temps de l'activité antérieure de l'utilisateur (p. ex. au cours de la dernière journée, 24-48 heures auparavant...8-30 jours auparavant).
Domain_name_top99	Nom de domaine du site dans lequel l'encart publicitaire est affiché. Limité aux 99 premiers domaines qui, au cours des expériences d'ingénierie de caractéristiques, ont révélé la capture de 95 % de tous les clics.
slot_visibility	Visibilité de l'encart publicitaire dans la page sur l'écran de l'utilisateur.
device	Identifiant du dispositif de l'utilisateur.

Nom de la caractéristique	Description de la caractéristique
fmt_device	Une caractéristique conçue comprenant une combinaison de format (fmt) d'offre et d'identifiant de dispositif d'utilisateur (dispositif).
ad_slot_key_top499	Un identifiant unique pour la combinaison : éditeur, encart publicitaire et page. Limitée aux premières 499 valeurs qui, au cours des expériences d'ingénierie de caractéristiques, ont révélé la capture de 97 % de tous les clics.
camp_type	Un identifiant catégorique de type de campagne associé à une offre (p. ex. texte + image, bannière d'affichage avec contenu dynamique, bannière d'affichage statique).
user_country_top3	Le pays à partir duquel l'utilisateur a accédé au site. Limité aux trois premiers pays qui, au cours des expériences d'ingénierie de caractéristiques, ont révélé la capture de plus de 99 % de tout le trafic. Il faut noter cependant que le nombre et l'identité des premiers pays sont spécifiques à un éditeur/bourse publicitaire qui peut être spécifique à une région où une langue.
offer_pos	Une valeur catégorique indiquant le placement d'une offre à l'intérieur d'un encart publicitaire.
browser	L'identifiant du navigateur d'un utilisateur (p. ex. Chrome, IE, Safari, etc.).

## REVENDEICATIONS

1. Un appareil informatique (102) qui, côté demande, implémente une plateforme, l'appareil informatique comprenant :

un processeur (104) ;

au moins un dispositif de mémoire (106, 110) accessible par le processeur ; et

une interface de communications de données (112) qui est associée de façon opératoire avec le processeur,

dans lequel le dispositif de mémoire contient un corps d'instructions de programme (114) incluant un modèle d'apprentissage automatique exécutable par le processeur et qui est configuré pour déterminer une probabilité d'interaction de l'utilisateur avec un élément de contenu, dans lequel le modèle d'apprentissage automatique est un modèle linéaire généralisé comprenant une pluralité de coefficients de modèle, dans lequel la pluralité des coefficients de modèle est stockée dans une structure de données de dictionnaire dans laquelle chaque entrée est définie par une clé et une valeur de coefficient, dans laquelle chaque clé comprend une représentation hachée d'une concaténation d'un nom de caractéristique et d'une valeur de caractéristique correspondante, le modèle ayant été formé en utilisant un ensemble de vecteurs de caractéristiques de formation enrichies et des balises correspondantes d'événements d'interaction qui sont dérivées d'un ensemble généré à partir d'enregistrements de données appariées qui concernent des événements de placement de contenu et des enregistrements concernant des événements d'interaction d'utilisateurs récupérés dans un entrepôt de données en ligne, sans lien explicite entre eux, dans lequel les évènements de placement et d'interaction surviennent au cours d'un laps de temps défini ;

le corps des instructions de programme incluant par ailleurs des instructions qui, lorsqu'elles sont exécutées par le processeur, amènent l'appareil informatique à mettre en œuvre un procédé de prédiction en temps réel comprenant les étapes suivantes :

la réception (602) d'informations via l'interface de communications de données concernant l'encart pour le placement de contenu en ligne et des informations concernant un utilisateur auquel le contenu de l'encart de placement de contenu en ligne sera affiché ;

le calcul (604) d'un vecteur de caractéristiques enrichies de probabilités, basé sur un élément de contenu sélectionné pour le placement à l'intérieur de l'encart de placement de contenu en ligne, les informations relatives à l'utilisateur et les informations relatives à l'encart de placement de contenu en ligne correspondant ;

l'exécution d'un modèle d'apprentissage automatique pour déterminer (606) une probabilité d'interaction de l'utilisateur avec l'élément de contenu sélectionné sur la base du vecteur de caractéristiques enrichies de probabilités.

2. L'appareil selon la revendication 1 dans lequel le modèle d'apprentissage automatique est un modèle de régression logistique.
  
3. L'appareil selon la revendication 2 dans laquelle les instructions de programme amènent l'appareil informatique à mettre en œuvre l'étape d'exécution du modèle d'apprentissage automatique en :
  - général, pour chaque valeur de caractéristique du vecteur de caractéristiques enrichies de probabilité, une clé correspondante ;
  - recupérant à partir de la structure de donnée de dictionnaire, pour chaque clé générée, une valeur de coefficient correspondante ; et
  - calculant, en utilisant le vecteur de caractéristiques enrichies de probabilité et les valeurs de coefficient récupérées, l'estimation de probabilité d'une interaction de l'utilisateur avec l'élément de contenu sélectionné.
  
4. L'appareil selon l'une quelconque des revendications 1 à 3 dans lequel :
  - l'encart de placement de contenu en ligne est un encart publicitaire ;
  - les informations relatives à l'encart publicitaire et les informations relatives à l'utilisateur auquel le contenu dans l'encart publicitaire sera affiché sont reçues

avec un message de demande d'enchère (210) transmis à partir d'un serveur de bourse publicitaire (122) ;

l'élément de contenu comprend au moins une enchère pour le placement à l'intérieur de l'encart publicitaire et

le corps des instructions de programme inclut par ailleurs des instructions qui, lorsqu'elles sont exécutées par le processeur, amènent l'appareil informatique à mettre en œuvre le procédé comprenant les autres étapes suivantes :

la transmission (612), au serveur de bourse publicitaire (122), d'un message de réponse à une enchère (214) en réponse au message de demande d'enchère (210) ;

répondant à la réception, en provenance du serveur de bourse publicitaire, d'une notification d'enchère réussie, la mise à jour (222) d'un entrepôt de données en ligne (166) avec des données d'événements de placement concernant le placement de l'élément de contenu ; et

répondant à la réception d'une notification (224) d'une interaction d'un utilisateur avec l'élément de contenu, la mise à jour (226) de l'entrepôt de données en ligne (166) avec les données d'événements d'interaction relatives à l'interaction de l'utilisateur avec l'élément de contenu.

5. Un appareil informatique (152) qui implémente la formation d'un modèle d'apprentissage automatique configuré pour estimer la probabilité d'interaction d'un utilisateur avec des éléments de contenu, dans lequel le modèle d'apprentissage automatique est un modèle linéaire généralisé comprenant une pluralité de coefficients de modèles, l'appareil informatique comprenant :

un processeur (154) ;

au moins un dispositif de mémoire (156, 160) accessible par le processeur ; et

un entrepôt de données (166) accessible par le processeur, dans lequel le dispositif de mémoire contient un corps d'instructions

de programme (164) incluant des instructions qui, lorsqu'elles sont exécutées par le processeur, amènent l'appareil informatique à de mettre en œuvre un procédé comprenant les étapes suivantes :

la génération d'une clé comprenant une représentation hachée d'un nom de caractéristique et d'une valeur de caractéristique correspondant au coefficient, et

le stockage, dans une structure de donnée de dictionnaire, d'une valeur de coefficient associée à la clé, dans lequel la structure de donnée de dictionnaire comprend un chiffage efficace du modèle d'apprentissage automatique.

l'accès (402) à l'entrepôt de données (166) pour récupérer des enregistrements concernant des événements de placement de contenu, et des enregistrements concernant des événements d'interaction d'un utilisateur, dans lequel surviennent les événements de placement et d'interaction au cours d'un laps de temps défini ;

l'appariement (404) des enregistrements d'événements et de placement de contenu avec des enregistrements d'événements d'interaction récupérés afin de générer un ensemble de données appariées comprenant une pluralité d'enregistrements, chaque enregistrement de l'ensemble de données appariées incluant un ensemble de valeurs de caractéristiques brutes dérivées d'un événement de placement de contenu avec une balise d'événement d'interaction indiquant si oui ou non un événement d'interaction correspondant à l'événement de placement de contenu est survenu ;

le calcul (406), à partir des valeurs de caractéristiques brutes, d'un ensemble correspondant de vecteurs de caractéristiques de formation ; et

la formation (408) du modèle d'apprentissage automatique en utilisant les vecteurs de caractéristiques de formation enrichies et les balises d'événements d'interaction correspondantes.

6. L'appareil selon la revendication 5 dans lequel le modèle d'apprentissage automatique est un modèle de régression logistique et les instructions de programme amènent l'appareil informatique à mettre en œuvre l'étape de formation du modèle d'apprentissage automatique en utilisant la régression logistique régularisée avec l'algorithme d'apprentissage proximal 'd'Amorce régularisée de suivi' (Follow-The-Regularised-Leader) (FTRL).

7. L'appareil selon les revendications 5 ou 6 dans lequel l'ensemble des instructions de programme inclut des instructions qui, lorsqu'elles sont exécutées par le processeur, amènent l'appareil informatique à exécuter de façon récurrente les étapes suivantes : l'accès (402) à l'entrepôt de données en ligne (166), l'appariement (404) des enregistrements récupérés d'événements de placement de contenu ainsi que les enregistrements récupérés d'événements d'interaction, le calcul (406) des vecteurs de caractéristiques de formation enrichies et la formation (408) du modèle d'apprentissage automatique pour actualiser le modèle d'apprentissage automatique.

8. Un procédé de prédiction en temps réel mis en œuvre informatiquement comprenant :

l'accès (402) à un entrepôt de données en ligne (166) pour récupérer des enregistrements concernant des événements de placement de contenu et des enregistrements concernant des événements d'interaction d'un utilisateur, n'ayant aucune lien explicite entre eux, dans lequel les événements de placement et d'interaction surviennent au cours d'un laps de temps défini ;

l'appariement (404) des enregistrements récupérés d'événements de placement de contenu avec des enregistrements récupérés d'événements d'interaction pour générer un ensemble de données appariées incluant un ensemble de valeurs brutes de caractéristiques dérivées d'un événement de placement de contenu avec une balise d'événements d'interaction indiquant si oui ou non un événement d'interaction correspondant avec l'événement de placement de contenu est survenu ;

le calcul (406), à partir des valeurs de caractéristiques brutes, d'un

ensemble correspondant de vecteurs de caractéristiques de formation enrichies ;

la formation (408) d'un modèle d'apprentissage automatique en utilisant les vecteurs de caractéristiques de formation enrichies et les balises d'événements d'interaction correspondantes, dans lequel le modèle d'apprentissage automatique est un modèle linéaire généralisé comprenant une pluralité de coefficients de modèle, dans lequel la pluralité des coefficients de modèle est stockée dans une structure de données de dictionnaire dans laquelle chaque entrée est définie par une clé et une valeur de coefficient, dans laquelle chaque clé comprend une représentation hachée d'une concaténation d'un nom de caractéristique et d'une valeur de caractéristique correspondante ;

la réception (602), au niveau d'un processeur (316) configuré pour exécuter le modèle d'apprentissage automatique, d'informations concernant l'encart de placement de contenu en ligne (148, 150) et des informations concernant un utilisateur auquel le contenu à l'intérieur de l'encart de placement de contenu en ligne sera affiché ;

le calcul (604), par le processeur, d'un vecteur de caractéristiques enrichies d'estimation basé sur un élément de contenu sélectionné pour le placement à l'intérieur de l'encart de placement de contenu en ligne, d'informations concernant l'utilisateur et d'informations concernant l'encart de placement de contenu en ligne ;

la détermination (606), par le processeur exécutant le modèle d'apprentissage automatique, d'une estimation de la probabilité d'une interaction de l'utilisateur avec l'élément de contenu sélectionné, basée sur le vecteur de caractéristiques enrichies d'estimation.

9. Le procédé selon la revendication 8 dans lequel :

l'encart de placement de contenu en ligne est un encart publicitaire ;

les informations relatives à l'encart publicitaire et les informations relatives à l'utilisateur auquel le contenu à l'intérieur de l'encart publicitaire sera affiché sont reçues avec un message de demande d'enchère (210) transmis par un serveur de bourse publicitaire (122) ; et

l'élément de contenu comprend au moins une enchère pour un placement à l'intérieur d'un encart publicitaire.

10. Le procédé selon la revendication 9 comprenant par ailleurs :  
la transmission (612), au serveur de bourse publicitaire (122) par le processeur (316), d'un message de réponse à l'enchère (214) en réponse au message de demande d'enchère (210) ;  
la réception, par le processeur en provenance du serveur de bourse publicitaire, d'une notification d'enchère réussie ; et  
la mise à jour (222), par le processeur, d'un entrepôt de données en ligne (166) avec les données de l'événement de placement de contenu concernant le placement de l'élément de contenu.

11. Le procédé selon la revendication 10 comprenant par ailleurs :  
la réception, par le processeur (316), d'une notification (224) d'interaction de l'utilisateur avec l'élément de contenu ; et  
la mise à jour (226) de l'entrepôt de données en ligne (166) avec les données d'événement d'interaction relatives à l'interaction de l'utilisateur avec l'élément de contenu.

12. Le procédé selon la revendication 11 dans lequel les étapes suivantes : l'accès (402) à l'entrepôt de données en ligne (166), l'appariement (404) des enregistrements récupérés d'événements de placement de contenu avec les enregistrements récupérés d'événements d'interaction, le calcul (406) de vecteurs de caractéristiques de formation enrichies et la formation (408) du modèle d'apprentissage automatique, sont exécutés de façon récurrente pour actualiser le modèle d'apprentissage automatique.

13. Un programme d'ordinateur comprenant un code de programme pour exécuter les étapes du procédé selon l'une quelconque des revendications 8 à 12 lorsque ledit programme fonctionne sur un ordinateur.

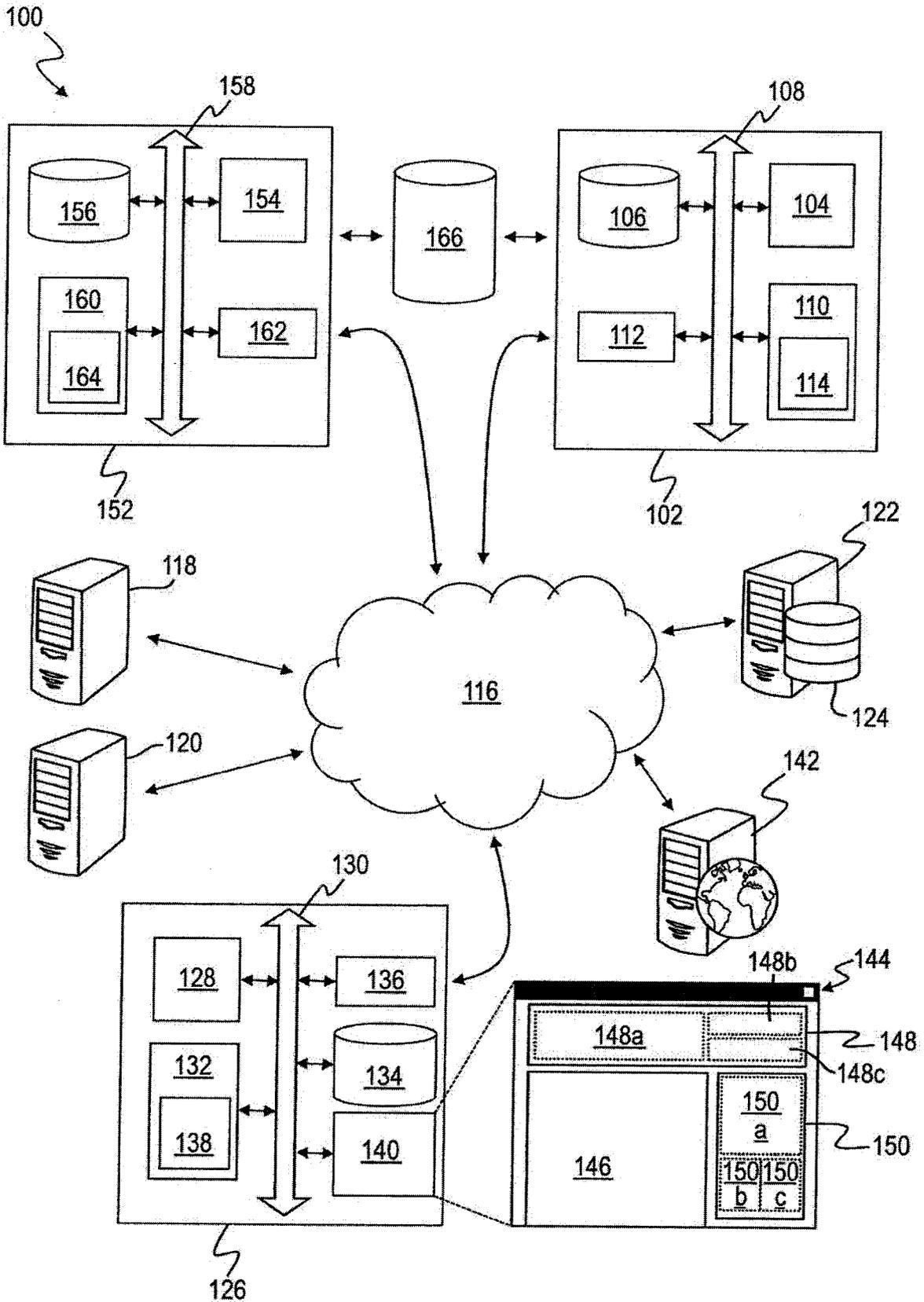


Figure 1

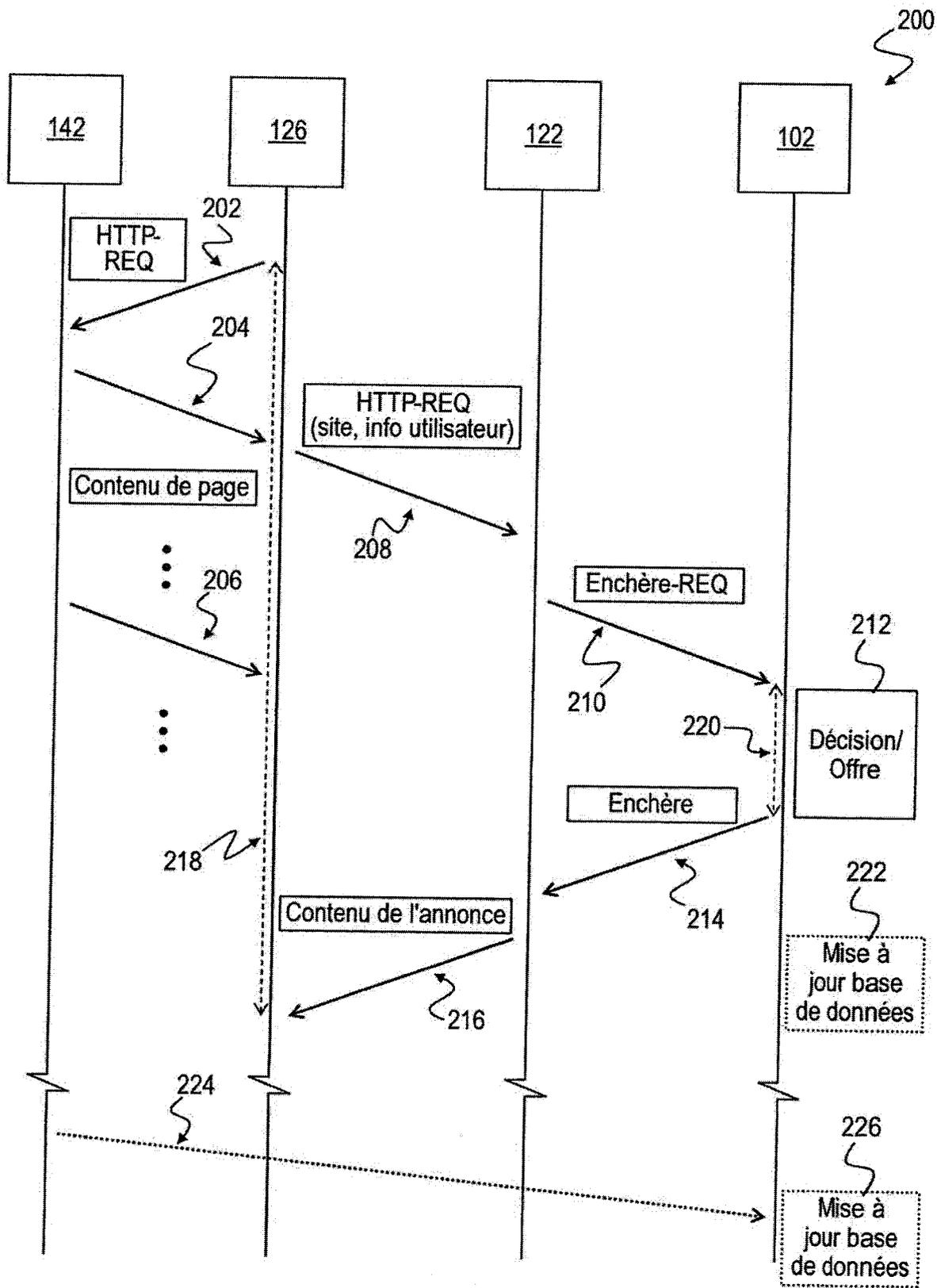


Figure 2

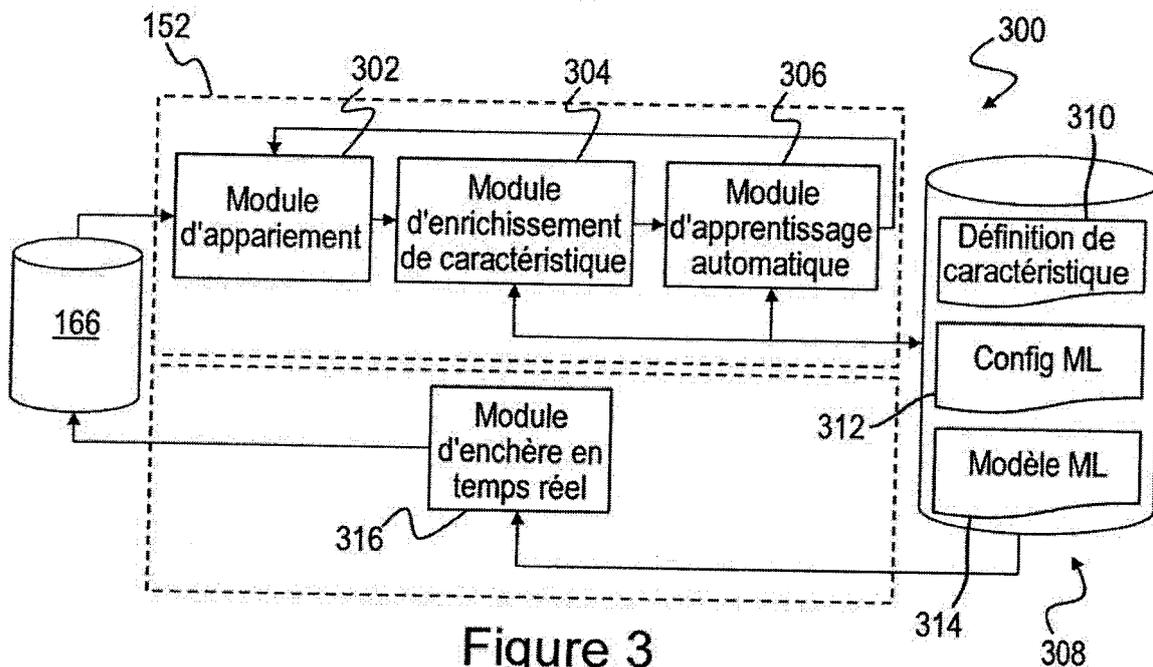


Figure 3

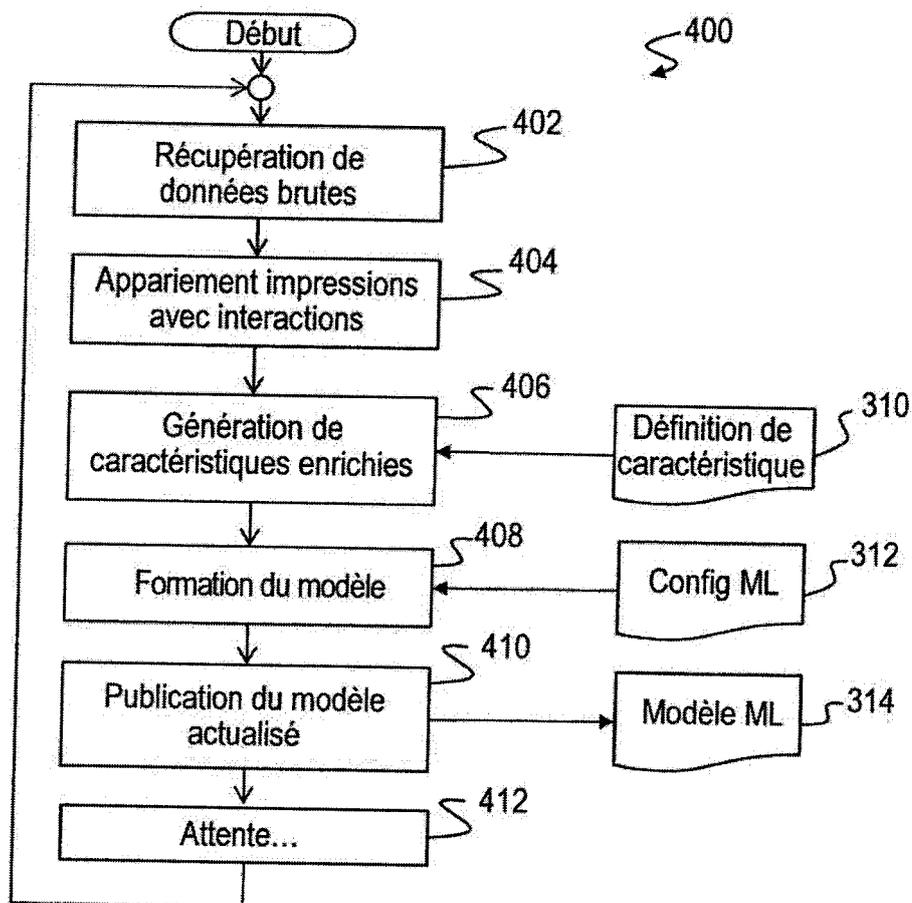


Figure 4

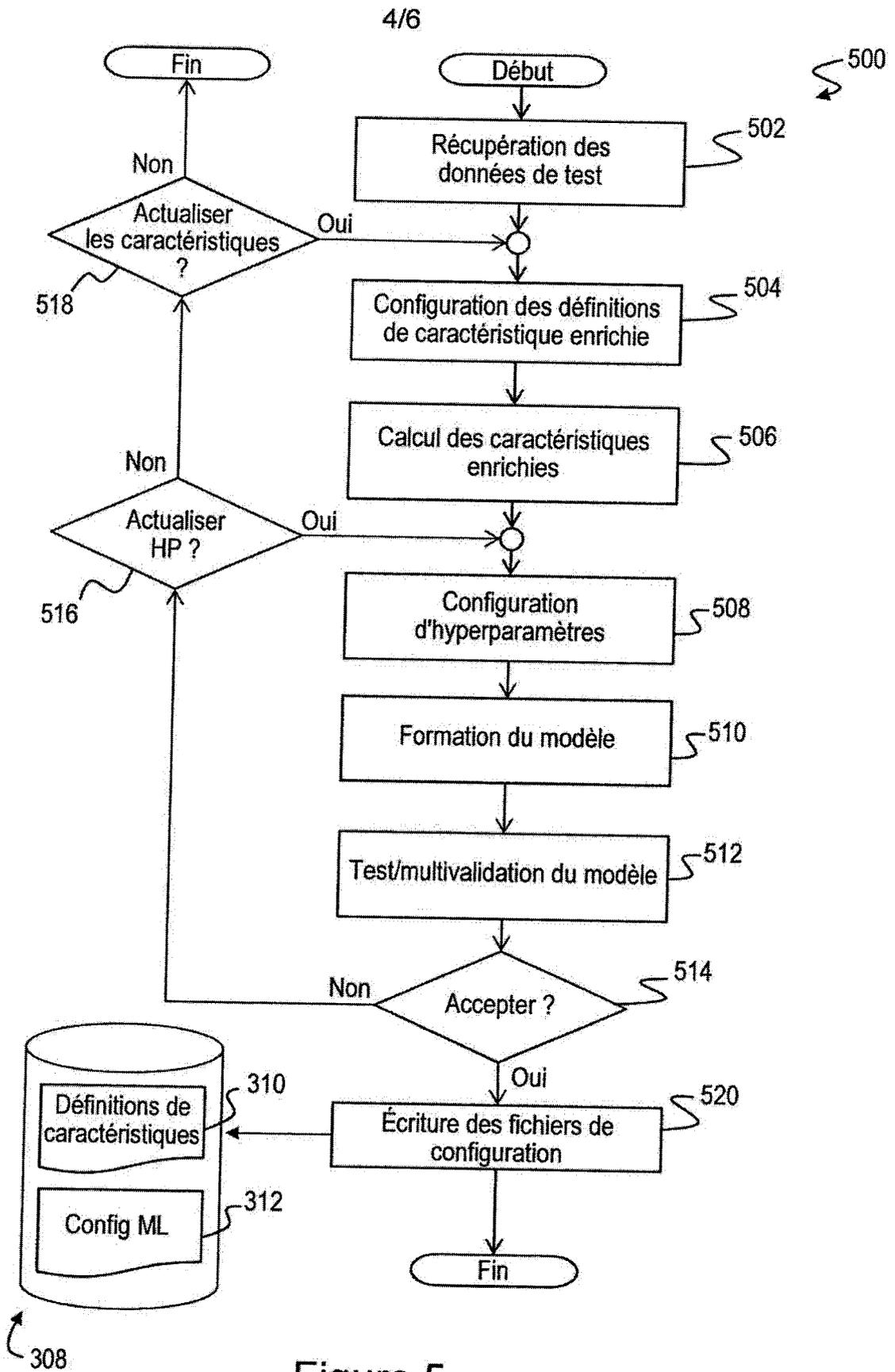


Figure 5

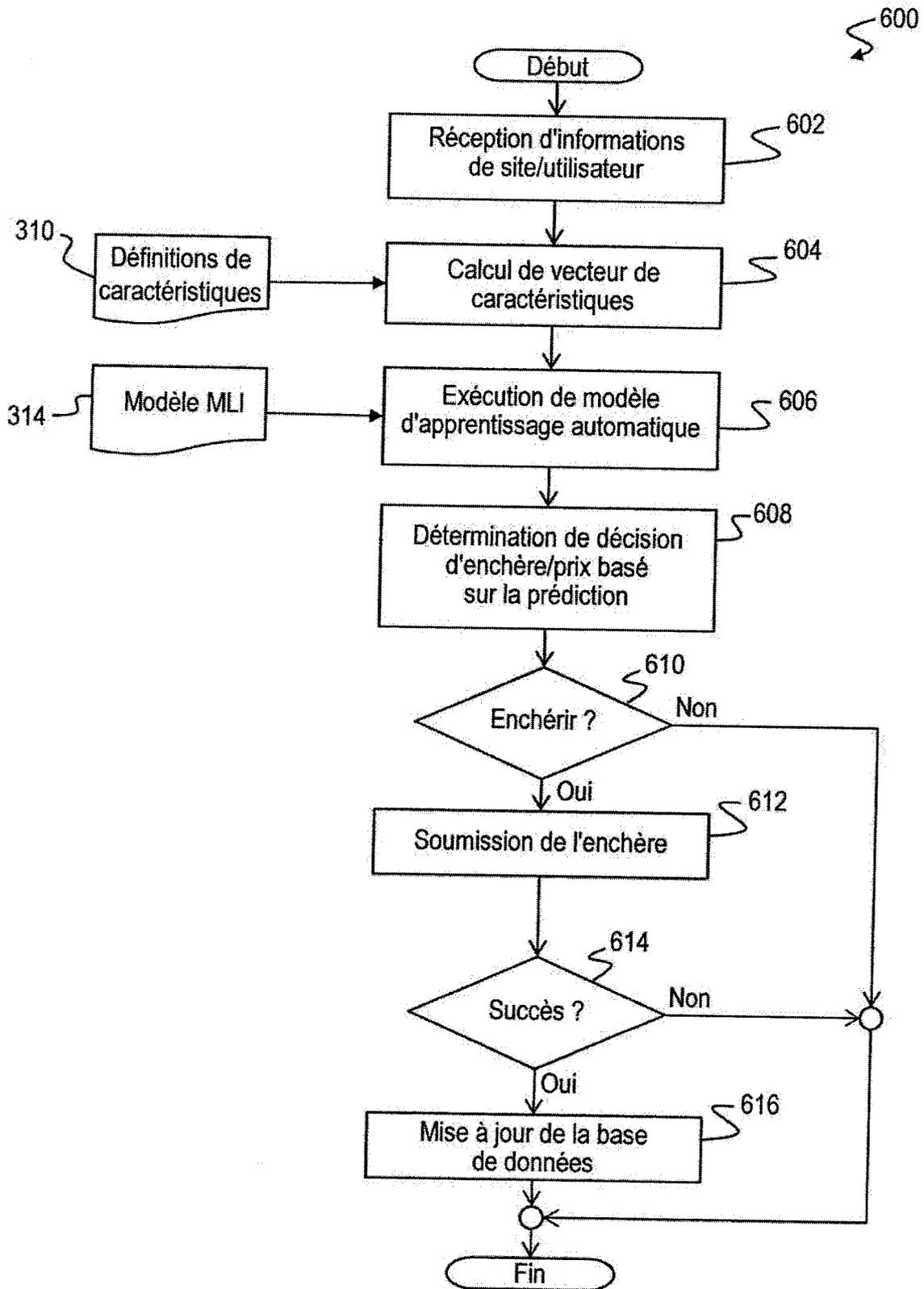


Figure 6

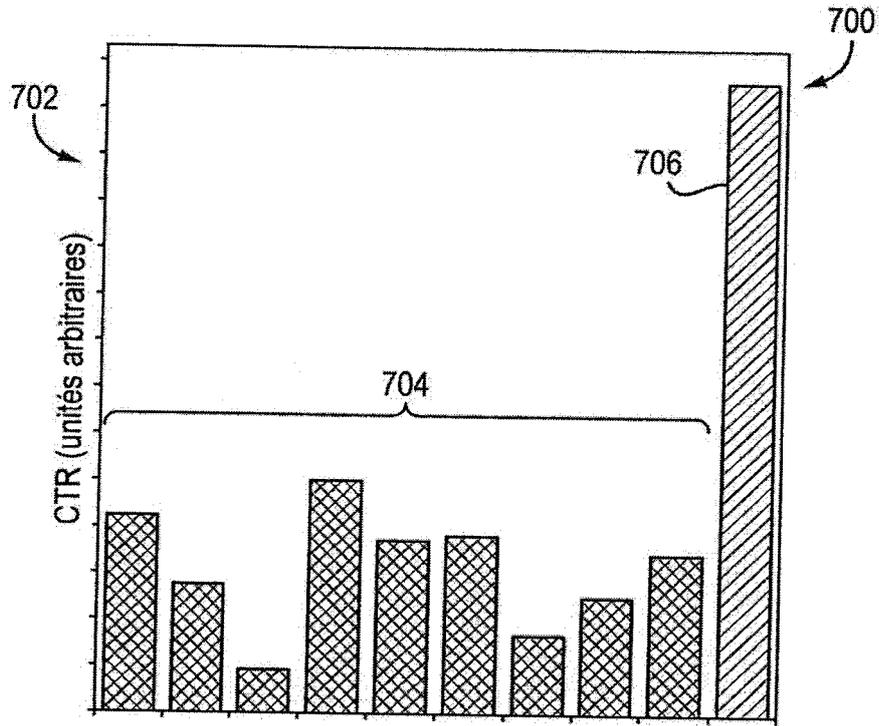


Figure 7(a)

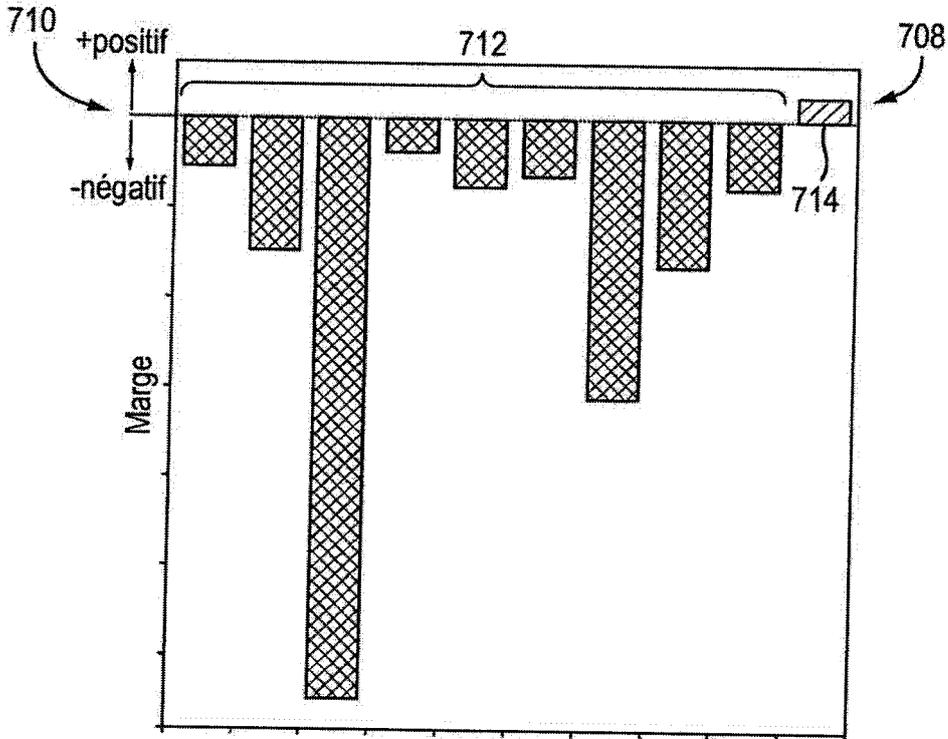


Figure 7(b)

# RAPPORT DE RECHERCHE

articles L.612-14, L.612-53 à 69 du code de la propriété intellectuelle

## OBJET DU RAPPORT DE RECHERCHE

L'I.N.P.I. annexe à chaque brevet un "RAPPORT DE RECHERCHE" citant les éléments de l'état de la technique qui peuvent être pris en considération pour apprécier la brevetabilité de l'invention, au sens des articles L. 611-11 (nouveau) et L. 611-14 (activité inventive) du code de la propriété intellectuelle. Ce rapport porte sur les revendications du brevet qui définissent l'objet de l'invention et délimitent l'étendue de la protection.

Après délivrance, l'I.N.P.I. peut, à la requête de toute personne intéressée, formuler un "AVIS DOCUMENTAIRE" sur la base des documents cités dans ce rapport de recherche et de tout autre document que le requérant souhaite voir prendre en considération.

## CONDITIONS D'ETABLISSEMENT DU PRESENT RAPPORT DE RECHERCHE

Le demandeur a présenté des observations en réponse au rapport de recherche préliminaire.

Le demandeur a maintenu les revendications.

Le demandeur a modifié les revendications.

Le demandeur a modifié la description pour en éliminer les éléments qui n'étaient plus en concordance avec les nouvelles revendications.

Les tiers ont présenté des observations après publication du rapport de recherche préliminaire.

Un rapport de recherche préliminaire complémentaire a été établi.

## DOCUMENTS CITES DANS LE PRESENT RAPPORT DE RECHERCHE

La répartition des documents entre les rubriques 1, 2 et 3 tient compte, le cas échéant, des revendications déposées en dernier lieu et/ou des observations présentées.

Les documents énumérés à la rubrique 1 ci-après sont susceptibles d'être pris en considération pour apprécier la brevetabilité de l'invention.

Les documents énumérés à la rubrique 2 ci-après illustrent l'arrière-plan technologique général.

Les documents énumérés à la rubrique 3 ci-après ont été cités en cours de procédure, mais leur pertinence dépend de la validité des priorités revendiquées.

Aucun document n'a été cité en cours de procédure.

**1. ELEMENTS DE L'ETAT DE LA TECHNIQUE SUSCEPTIBLES D'ETRE PRIS EN CONSIDERATION POUR APPRECIER LA BREVETABILITE DE L'INVENTION**

US 2005/131762 A1 (BHARAT KRISHNA [US] ET AL) 16 juin 2005 (2005-06-16)

US 2013/254787 A1 (COX EARL [US] ET AL) 26 septembre 2013 (2013-09-26)

EP 1 223 757 A2 (METABYTE NETWORKS INC [US]) 17 juillet 2002 (2002-07-17)

CN 101 484 890 A (IN V IDI TECHNOLOGIES CORP [US]) 15 juillet 2009 (2009-07-15)

JP 2016 062411 A (YAHOO JAPAN CORP) 25 avril 2016 (2016-04-25)

**2. ELEMENTS DE L'ETAT DE LA TECHNIQUE ILLUSTRANT L'ARRIERE-PLAN TECHNOLOGIQUE GENERAL**

NEANT

**3. ELEMENTS DE L'ETAT DE LA TECHNIQUE DONT LA PERTINENCE DEPEND DE LA VALIDITE DES PRIORITES**

NEANT