

Assistance à la découverte de connaissances contextuelles à partir de l'analyse des traces

Assitan Traoré^{1,2}, Alain Mille¹ et Hélène Tattegrain²

¹IFSTTAR, LESCOT

{assitan.traore, helene.tattegrain}@ifsttar.fr

<http://www.ifsttar.fr>

²Université Lyon 1, LIRIS, CNRS UMR 5205

{assitan.traore, alain.mille}@univ-lyon1.fr

<http://liris.cnrs.fr>

Résumé : L'observation d'une activité pour comprendre des comportements particuliers nécessite de discerner ce qui relève ou non du contexte. En intelligence artificielle la notion de contexte est une approche modale du raisonnement et plusieurs études ont été faites pour proposer des modèles génériques de représentation des connaissances *contextuelles*. La question de la découverte interactive des connaissances contextuelles et de leur inscription dans un modèle générique n'est toutefois que rarement examinée dans la littérature. Cette communication propose une approche exploitant le potentiel de la représentation de l'activité par ses traces modélisées d'interaction. Elle consiste à assister la découverte du contexte explicatif par une approche interactive associant l'analyse des données, les connaissances de l'analyste et leur mise en situation dans les traces modélisées. Cette approche interactive facilite l'élicitation de ce qui est contexte ou non. De plus, nous montrons qu'il est alors possible de représenter le contexte découvert sous une forme générique telle que proposée dans la littérature spécialisée. La méthode a été implémentée et a permis la découverte de contextes explicatifs pour la consommation en carburant lors de l'activité de conduite automobile. Les expérimentations, les données et les analyses ont été menées à l'IFSTTAR dans des conditions de conduite réelles.

Mots-clés : Découverte interactive de connaissances, contexte, traces modélisées, transport, consommation en carburant.

1 Introduction

La question de l'ingénierie des connaissances *contextuelles* se pose de manière aigüe dans le cadre de l'analyse d'activités complexes comme les comportements de conduite automobile, se déroulant dans des environnements variés, avec de fortes dynamiques temporelles et spatiales. C'est alors le *contexte* qui permet d'expliquer les différences entre divers comportements pour réaliser une activité similaire. Ce cadre d'analyse de l'activité de conduite centré sur l'explication de la consommation de carburant a été utilisé pour étudier et proposer une méthode de découverte interactive de contexte à partir d'une analyse de l'activité tracée selon l'approche des *traces modélisées*. L'article précise tout d'abord le cadre applicatif du travail avant de réaliser l'état de l'art sur la notion de contexte, de sa modélisation et de sa découverte. L'approche est ensuite décrite avec ses différentes étapes de la préparation des données à la découverte des éléments de contexte « candidats ». Cette approche a été appliquée sur un jeu de données dans le domaine de la conduite automobile, et la première analyse menée avec l'aide de cette approche est décrite. Cette analyse montre des résultats très encourageants, démontrant le potentiel de cette approche de découverte interactive d'éléments de contexte.

2 Cadre du travail de recherche

Ce travail est effectué en collaboration entre le LIRIS et l'IFSTTAR dans le but d'étudier et d'expérimenter une méthodologie de découverte de connaissances contextuelles en l'appliquant à l'explication d'un critère de classification de segments temporels d'une activité.

2.1 Enjeux sociaux pour la recherche dans les transports

Le phénomène du réchauffement climatique est attribué, entre autres, aux gaz à effet de serre produits par les activités humaines l'industrie, le transport, le résidentiel-tertiaire et l'agriculture. L'un des objectifs principaux de l'IFSTTAR est de réduire l'impact environnemental du transport. Les études environnementales montrent que le transport est l'un des principaux responsables du réchauffement climatique (RAC-F, 2007). Plusieurs études ont été menées pour remédier à ce problème environnemental. Elles ont vulgarisé en 1999 le terme d'Eco-conduite avec des systèmes d'assistance de conduite comme: ISA, LAVIA (Rakotonirainy et al., 2011). Des études récentes sur les impacts des systèmes d'Eco-conduite et d'assistance à la conduite sur la consommation montrent que ces systèmes sont moins efficaces dès que la vitesse limite est supérieure à 80km/h (par exemple sur l'autoroute) et que les cours d'Eco-conduite ne réduisent la consommation que pendant une durée courte après l'apprentissage du conducteur (Beusen et al., 2009). D'autre part, cet impact environnemental est amplifié par l'augmentation constante du parc automobile, avec en France, 5% d'augmentation sur les 10 dernières années (MEDDE, 2014). D'autres pistes sont ouvertes pour proposer des systèmes d'assistance capables de réduire la consommation de carburant de façon fiable et durable dans des contextes variés.

2.2 Présentation de la question de recherche

« Le contexte est toute information qui peut être utilisée pour caractériser la situation d'une entité. Une entité peut être une personne, un endroit ou un objet que l'on considère pertinent dans l'interaction entre un utilisateur et une application, y compris l'utilisateur et l'application eux-mêmes » (Dey, 2001). D'après cette définition du contexte de Dey, pour caractériser une activité il faut tenir compte des informations relatives aux entités (les éléments en contexte) qui sont impliquées dans la réalisation de cette activité. Mais ces informations sont inconnues au début de l'activité et ne sont définies que pendant la réalisation de cette dernière. Par exemple, pour modéliser les comportements de conduite réduisant la consommation en carburant, il faut tenir compte des entités impliquées dans l'activité de conduite comme le véhicule, le conducteur et l'environnement de conduite et déterminer les valeurs prises par ces entités lorsque la consommation en carburant est faible ou forte pendant la conduite. Les informations prises par les entités de l'activité correspondent aux facteurs dont dépend la consommation en carburant comme le type du véhicule, l'état du véhicule, le mode de conduite du conducteur, l'infrastructure, le trafic. Ces facteurs décrivent les entités de l'activité de conduite et ils permettent d'avoir une définition précise de l'activité à un moment précis ou sur un segment temporel donné. Ils pourront être utilisés pour expliquer des comportements précis ou des phénomènes observés pendant l'activité. Les éléments contexte sont difficiles à définir en amont car ils sont dynamiques. Pour les définir nous proposons un processus de découverte à partir de l'analyse des traces (Georgeon et al., 2012 ; Mathern, 2012) associée à des méthodes de fouille de données et à l'expertise de l'analyste.

2.3 Présentation des données utilisées

Les données utilisées dans ce travail ont été collectées lors de l'activité de conduite de plusieurs participants pendant plusieurs mois. Ces données ont été recueillies en conduite *naturelle* avec des véhicules équipés de capteurs (bus CAN, radar, GPS et cameras).

3 Etat de l'art de la notion de contexte et de l'ingénierie des connaissances contextuelles

3.1 La notion de contexte en intelligence artificielle : une histoire ancienne qui pose la question d'une ingénierie spécifique des connaissances contextuelles

La notion de contexte a été étudiée très tôt en intelligence artificielle par John McCarthy, (McCarthy, 1993). Dans cet article, la relation de base $ist(c,p)$ établit que la proposition p est vraie dans le contexte c . Ceci permet d'étendre ou de restreindre un raisonnement en fonction d'un contexte, qu'il faut alors établir. Le contexte porte un nom « c » et est défini en extension par les variables qui sont contenues dans les propositions qui permettent de produire une inférence valide dans ce contexte. La liste des variables à mobiliser dans ce contexte constituent, d'une certaine façon, le contexte en extension. Comment identifier ces variables ? Ces travaux ont été rapidement approfondis et comparés et le lecteur intéressé trouvera une étude comparative très bien documentée dans (Akman et al., 1996). Comment savoir à quelles hypothèses répondent tels ou tels contextes ? En pratique, il est ressorti rapidement que la définition d'un contexte était problématique et soulevait aussi bien des questions de complexité de calcul que de difficultés à les définir. La complexité de l'ingénierie de cette modélisation pour la représentation des connaissances dans le système CYC (Lenat, 1998) a imposé une méthode de réduction de complexité. En effet, une grande variété de combinaisons de valorisations de variables permet de définir un contexte, et chaque contexte doit être « affirmé » en examinant la validité des hypothèses de validité des règles que ses variables peuvent (ou non) déclencher. Pour résoudre cette question, Lenat, dès 1998, propose de décrire « presque » n'importe quel contexte selon « seulement » 12 dimensions qui permettraient de décrire 99 % des contextes qu'il a imaginés. Ces dimensions sont (nous reprenons la liste en anglais telle qu'elle a été publiée):

- *Absolute Time: a particular time interval in which events occur*
- *Type Of Time: a non-absolute type of time period, such as "just after eating"*
- *Absolute Place: a particular location where events occur, such as "Paris"*
- *Type Of Place: a non-absolute type of place, such as "in bed"*
- *Culture: linguistic, religious, ethnic, age-group, wealth, etc. of typical actors*
- *Sophistication/Security: who already knows this, who could learn it, etc.*
- *Topic/Usage: drilling down into aspects and applications – not subsets*
- *Granularity: phenomena and details which are (and are not) ignored*
- *Modality/Disposition/Epistemology: who wants/believes this content to be true?*
- *Argument-Preference: local rules for how to resolve pro-con argument disputes*
- *Justification: are things in this context generally proven, observed, on faith...*
- *Let's: local bindings of variables etc. that hold true in that context.*

Ces dimensions sont surtout descriptives et chaque définition possède des raffinements importants, en particulier pour la gestion du temps. En pratique, avant de lancer une inférence, il faut définir un « motif contextuel » (un intervalle, une zone géographique, une contrainte sur telle ou telle autre dimension) et sélectionner les règles d'inférence qui sont valides pour ce motif de contexte, ce qui laisse supposer que toute règle est annotée par le motif contextuel pour lequel elle est valide.

(Brézillon, 2000) fait le point des connaissances sur la notion de contexte en y associant les notions de révision, de raisonnement hypothétique, de raisonnement analogique également très étudiés en Intelligence Artificielle. Avec Fausto Giunchiglia (Giunchiglia et al., 1996 ; Bouquet et al., 2001) Patrick Brézillon anime la série des conférences CONTEXT¹ depuis 2007. Cette conférence a permis d'élargir considérablement l'étude de la notion de contexte à d'autres champs de recherche que l'intelligence artificielle. La section suivante se fonde principalement sur ces derniers travaux pour en faire un rapide état de l'art, en particulier sous

¹ <http://cyprusconferences.org/context2015/>

l'angle de la modélisation, en l'analysant avec le prisme de l'ingénierie des connaissances pour l'appliquer au cas particulier de l'observation d'une activité.

3.2 La modélisation des connaissances contextuelles lors de l'observation d'une activité

La définition la plus reprise du contexte est celle de (Dey, 2001). Pour Dey une information est considérée comme du contexte si cette information permet de caractériser la *situation* d'une entité engagée dans une activité. Les entités engagées dans une activité sont les éléments qui expliquent spécifiquement une séquence particulière dans l'activité en question.

*Par exemple dans un accident de la route le fait que le conducteur ait consommé récemment de l'alcool ou pas est une information relevant du contexte. En effet, elle permet de savoir si le conducteur était dans un état normal pour conduire ou non. Cette information caractérise la **situation**² de l'entité « conducteur » engagée dans l'activité de conduite au moment de l'accident.*

Cette définition est générale car, quel que soit le domaine, elle permet de déterminer les informations relevant du contexte. Si cette définition est retenue, la question de la modélisation reste entière. (Dey et al., 1999) proposent une plateforme de gestion des connaissances contextuelles en suivant cette approche méthodologique. (Gandon et al., 2004) proposent une variante de cette approche garantissant une plus grande interopérabilité. Ces travaux montrent que l'on saurait gérer les connaissances de contexte, y compris pour les exploiter pour des applications tierces. La question de leur découverte, posée comme difficile par tous les auteurs, n'apparaît toutefois pas dans les plateformes proposées.

En 2002, (Rey et al., 2002) définissent le contexte comme un espace infini d'informations évolutives, qui ne sont pas connues en avance. Ce caractère dynamique du contexte rend sa modélisation très difficile. Pour réduire cette complexité de représentation du contexte d'une activité, en 2005, (Bazire et al., 2005) proposent une décomposition générique d'une activité selon les principales entités qu'elle mobilise. Ce modèle représente les entités ou composants d'une activité et les relations qui peuvent exister entre eux pour représenter le contexte de l'activité. Le modèle générique de (Bazire et al., 2005) permet de catégoriser les informations candidates au contexte des entités d'une activité mais elle ne dit rien sur la manière de découvrir ces informations. Il s'agit d'une question d'ingénierie de la connaissance qui associe l'acquisition des connaissances auprès des experts et les méthodes cherchant à découvrir des connaissances dans les données caractéristiques de l'activité mobilisant ces connaissances.

Dans le cadre de ce travail, nous retenons cette représentation générique telle qu'elle est aujourd'hui reprise par la communauté CONTEXT, sans étudier sa validité, ce qui est, hors de notre question de recherche. Nous nous intéressons par contre à l'étude d'une ingénierie des connaissances qui soit adaptée à ce type de connaissance *contextuelle*. Dans un travail se référant à la question de l'ingénierie des connaissances contextuelles, (Castelli et al., 2008), les auteurs considèrent la manière de « produire » les informations contextuelles à partir de capteurs RFID, sans aborder la question de la construction du sens qui devra ensuite avoir lieu pour exploiter cette information estampillée « contexte ». Un travail plus orienté sur la découverte de connaissances contextuelles, (Ngoc et al., 2005) dans le cadre d'un middleware robotique, repose sur l'association de plusieurs méthodes d'apprentissage automatique à partir d'un jeu de capteurs et actionneurs avec une mise en correspondance avec une ontologie contextuelle pour l'interprétation. Les auteurs ne s'intéressent pas à la construction de l'ontologie. Un travail de même nature (Milea et al., 2008), mais plus général encore, propose une manière d'intégrer la médiation contextuelle dans un processus de découverte de connaissance. La notion de contexte y est bien définie et les connaissances contextuelles permettent de multiplier les points de vue sur les patterns découverts. La connaissance découverte est en relation avec le contexte qui lui est attribué, mais cette connaissance

² Travaux argumentant sur la nécessité de formaliser les contextes revendiquent une approche « située » de la cognition, telle que Clancey l'a théorisé par exemple (McCarthy, 1993).

contextuelle doit venir de l'expert médiateur de la découverte de connaissance. Plusieurs travaux s'intéressent à la représentation ontologique des contextes avec par exemple le langage CoOL (Strang et al., 2003), tandis que des travaux européens (Mullins et al., 2008) se sont efforcés d'exploiter ces possibilités dans le cadre de l'interopérabilité ubiquitaire.

Nous n'avons pas repéré de travaux spécifiquement dédiés à la question de la découverte de connaissances contextuelles. Il est difficile pour un expert de nommer et qualifier les contextes explicatifs pour chaque situation observable dans une activité comme le montre le champ de recherche de l'IFSTTAR qui cherche à élucider les contextes explicatifs de comportement de consommation de carburant. L'observation instrumentée de l'activité produit des données suffisamment riches et nombreuses pour potentiellement expliquer les situations à étudier. La découverte de connaissances à partir de données, utilise des méthodes numériques (statistiques surtout) ou symboliques (algorithmes de fouille surtout) pour l'identification de motifs, avec une approche supervisée ou non, à partir de données collectées dans le monde « réel » observé. Cette approche mobilise différemment les connaissances de l'expert analyste qui est sollicité aussi bien pour la préparation des données que pour l'interprétation des motifs découverts. (Fayyad et al., 1996a, 1996b) proposent la définition suivante : « La découverte de connaissances dans des bases de données est le processus non trivial d'identification, dans les données, de motifs valides, nouveaux potentiellement utiles et ultimement compréhensibles. » et ils proposent également le cycle de ce processus de découverte de connaissance. Lorsque les données ne sont pas « données », mais collectées à la suite de l'observation d'une activité produisant des traces de cette activité, nous parlerons alors d'ingénierie des connaissances *tracées*, en considérant que le choix de ce qui est observé est déjà le fruit d'une connaissance explicite qu'il convient de modéliser dès la collecte. Dans le cadre de cet article, nous nous appuyons sur cette approche pour étudier les possibilités offertes pour assister l'ingénierie des connaissances contextuelles. Nous rappelons les travaux de notre équipe de recherche et l'environnement mobilisé pour cette étude.

3.3 Ingénierie des connaissances à partir de traces d'observation

L'ingénierie des connaissances à partir de traces d'observation est une forme d'ingénierie de découverte *dynamique* des connaissances, dans le sens où la connaissance s'établit dynamiquement et que cette dynamique est explicitement gardée comme explicative de telle ou telle interprétation. Les connaissances manipulées réfèrent à des observations situées dans le temps (obsels : observed elements). Le concept de trace modélisée a été développé par l'équipe SILEX du laboratoire LIRIS dans le but de construire des connaissances à partir de l'observation des interactions observables dans une activité (Laflaquiere et al., 2008 ; Settouti et al., 2011). Le modèle d'une trace en fournit une certaine sémantique, le « vocabulaire » de la trace (les types d'obsels observables et les types de relations observables) et les types des attributs de chaque type d'obsel. Une M-Trace est donc déjà une certaine interprétation explicite de l'observation. Chaque Obsel est situé temporellement selon le modèle de représentation du temps de la trace (représentation temporelle ou séquentielle, temps absolu ou relatif, unités de temps utilisées, etc.). Ce modèle explicite est utilisé aussi bien pour documenter la trace auprès des utilisateurs que pour permettre des calculs inférentiels appelés transformations de trace. La trace *première* issue de la collecte peut en effet être transformée par des opérateurs de transformation construisant une trace transformée si un motif d'obsels est reconnu dans cette trace première, selon une interprétation particulière et explicite. Le même processus de transformation s'applique à une trace transformée pour raffiner un raisonnement. Des raisonnements différents (différents points de vue) permettent de construire des interprétations différentes à partir de la même observation. D'un point de vue épistémologique, cela revient à interpréter différemment le même jeu de données. Une M-Trace contient ainsi, non seulement des informations relatives à ce qui a été observé, mais aussi sur la manière d'interpréter ces éléments observés. C'est cette capacité à exprimer des points de vue différents que nous souhaitons exploiter pour mettre en évidence des contextes différents à définir pour expliquer convenablement tel ou tel motif comportemental. Nous présentons ci-dessous l'état actuel de cette étude et les premiers résultats que nous obtenons.

4 Proposition d'un processus d'assistance à la découverte de connaissances contextuelles à partir de l'expérience tracée

L'objectif de cette méthode est d'identifier les facteurs contextuels pouvant expliquer un critère sur des données d'activité temporelle. Le contexte étant spécifique à la tâche en cours (sous partie de l'activité globale), il est important de travailler sur des segments homogènes de l'activité pour identifier quels sont les composants contexte dont les variations vont expliquer celles du critère à explorer. Nous cherchons à construire une méthode de découverte de connaissance interactive et itérative basée sur l'observation et sur des connaissances du domaine de l'analyste. Elle doit être *interactive* car c'est l'analyste qui pilote chaque étape et doit être *itérative* car l'analyste peut revenir en arrière à n'importe quelle étape. Dans cette approche, l'analyste a un rôle très important car c'est lui qui guide l'analyse en fonction de ses connaissances du domaine et de l'activité. C'est lui qui propose les transformations sur une M-Trace (source) à faire pour « voir » l'activité selon « son » interprétation (M-Trace transformée). Ces transformations sont conservées dans la base de traces modélisées et la trace transformée obtenue peut alors être explicitée facilement et confrontée à d'autres interprétations également représentées par des transformations. Les observations collectées constituent les M-Traces premières et les M-Traces transformées constituent les différentes façons d'interpréter les choses selon telle ou telle expertise. La découverte est *supervisée* et les séquences à caractériser sont *étiquetées* par l'analyste qui y « reconnaît » quelque chose qui fait sens. Cinq étapes sont proposées pour ce processus dans les paragraphes suivants.

4.1 Définition des composants génériques du contexte

Cette étape nous permet à partir du modèle des composants contexte de (Bazire & Brézillon, 2005) d'identifier « toutes » les variables possibles « candidates » au contexte lors de l'analyse d'une activité en fonction des objectifs ou des besoins d'analyse. Le modèle des composants contexte est constitué des contextes des entités qui interviennent lors de la réalisation d'une activité. Par suite, le contexte d'une activité est défini par l'ensemble des contextes de chaque entité de cette activité qui sont : le contexte de l'utilisateur « Cu », le contexte de l'item (système, objet, application, etc.) « Ci », et le contexte environnemental « Ce ». Toutes les activités possèdent ces entités car une activité se traduit par l'interaction entre un utilisateur et un objet ou une application dans l'environnement de la tâche.

Le contexte de l'observateur « Co » permet de définir le contexte de l'analyse comme les objectifs et le périmètre de l'analyse d'une activité. Les informations relatives aux entités impliquées dans une activité peuvent être utilisées pour caractériser cette activité ou un segment particulier de cette dernière. A ce niveau de l'analyse il s'agit d'informations *potentielles* de contexte de l'activité analysée. Pour identifier celles qui sont réellement explicatives, nous utiliserons donc les techniques de l'ingénierie des connaissances tracées.

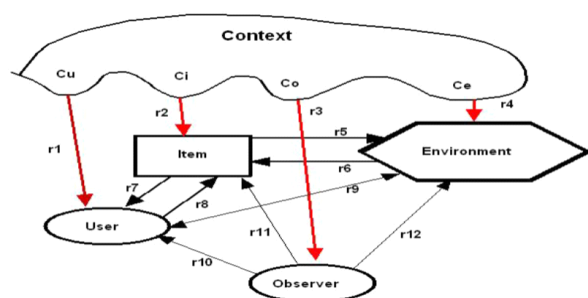


FIGURE 1 – Modèle des composants contexte (Bazire et al., 2005).

4.2 Collecte et préparation des données

Cette étape a pour but de faire la sélection et la contextualisation des données de l'activité en fonction des objectifs d'analyse. Elle mobilise les techniques de préparation de données telles que la gestion des valeurs manquantes, la création de nouvelles variables, la transformation des variables. Puis elles seront contextualisées en ajoutant des informations

relatives aux différents composants contexte définis dans le paragraphe 4.1. Il s'agit donc d'associer pour chaque segment d'activité sélectionné les variables potentielles explicatives du contexte. Ces variables peuvent être utilisées pour caractériser les contextes spécifiques aux différents comportements observés de l'activité analysée. Par exemple, pour le composant « Ce » de la partie 4.1, il faut collecter toutes les informations relatives à la localisation de l'activité et les contraintes associées. La contextualisation des données consiste donc à ajouter les informations permettant de définir, de caractériser de façon précise le contexte de chacune des entités de l'activité. Ces informations peuvent être collectées automatiquement à partir de capteurs (ex. le type de route) ou par codage manuel de l'activité (ex. le trafic, la météo).

4.3 Construction de la base de connaissances (les modèles des M-Traces)

La base de connaissances est composée de règles de reformulations pour décrire une activité ou une situation de l'activité comme une succession d'événements pertinents. Cette étape peut être très rapide si l'analyste maîtrise suffisamment le domaine pour créer directement ces règles lors de l'analyse tracée de l'activité. L'élaboration même de ces reformulations est très utile à l'analyste pour préciser et formaliser ses connaissances. Par exemple, dans cet article, cette étape a été utilisée pour segmenter l'activité de conduite en fonction des situations homogènes de conduite comme AllerToutDroit, TournerADroite etc.

4.4 Processus d'assistance à l'analyse à la découverte de connaissances contextuelles

Le processus d'assistance à la découverte du contexte lors de l'analyse d'une activité est effectué à partir de l'analyse des traces. Tracer une activité consiste à la décrire avec les aspects pertinents (*obsels*) à partir d'observations et des connaissances de l'analyse. Dans l'étape 4.3, les données ont été préparées. Avec cette base de données cible, la M-Trace première collectée T_0 est créée. A partir de T_0 , l'objectif est de décrire une situation comportementale d'une activité et d'en expliciter le contexte. Dans un premier temps, les premières transformations T_n vont permettre de créer les différents segments temporels de l'activité correspondant aux différentes tâches qui constituent l'activité analysée. Puis la transformation T_{sh} permet de sélectionner dans ces segments, ceux qui correspondent à la tâche de l'activité que l'on souhaite analyser (ex. segments S1 dans la

FIGURE 2).

Ces transformations sont faites grâce aux règles élaborées lors de l'étape 4.3 et complétées par l'expertise de l'analyste. La transformation T_{crit} , permet alors d'étiqueter les segments en fonction des valeurs du critère à expliquer (S1-- et S1++ dans la

FIGURE 2). Pour identifier parmi les variables contextuelles candidates, celles qui font vraiment partie du contexte, il est tenté, à partir de l'ensemble des segments homogènes de l'activité, des transformations $T_{contexte1}$, $T_{contexte2}$ en fonction des valeurs du critère à expliquer. Ces transformations sont effectuées en utilisant uniquement les variables candidates du contexte définies dans l'étape 4.1.

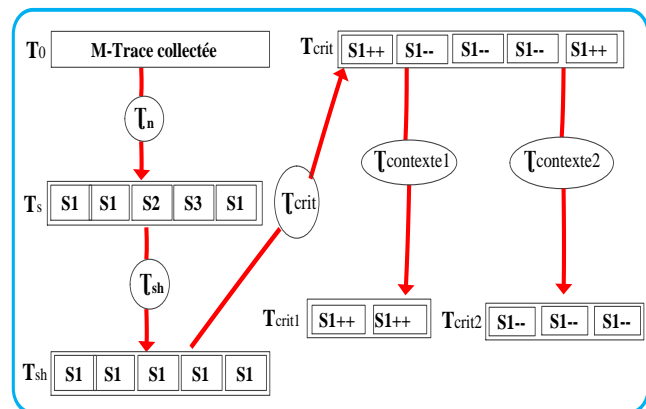


FIGURE 2 – Processus d'assistance à la découverte de connaissances contextuelles.

$T_{contexte1}$, $T_{contexte2}$ permettent de construire les M-traces T_{crit1} et T_{crit2} . Le contexte explicatif des critères $crit1$ et $crit2$ est défini par les variables candidates du contexte de l'activité qui ont été retenues respectivement dans les transformations $T_{contexte1}$, $T_{contexte2}$.

4.5 Evaluation et validation des variables de contexte à retenir

L'étape d'évaluation consiste à appliquer les connaissances découvertes sur de nouveaux jeux de données afin d'en déterminer la stabilité et la validité. Cela consiste à voir si les connaissances découvertes sont suffisamment stables sur de nouveaux jeux de données pour permettre une modélisation d'un modèle d'assistance contextuel à partir de ces connaissances. Cette validation est faite à l'aide des transformations considérées comme satisfaisantes, en les appliquant sur les M-Traces (observation des transformations issues de la partie 4.4 sur de nouveaux jeux de données) du corpus d'observation complet. Si le taux de réussite des transformations est supérieur à un certain taux, les variables retenues dans ces transformations sont intégrées dans la représentation du contexte à prendre en compte.

5 Application à la découverte du contexte dans le domaine des transports

Dans notre cas d'analyse, le processus d'assistance de découverte du contexte est appliqué au domaine du transport dans le but d'expliquer la consommation en carburant selon des comportements de conduite automobile. Pour cela, nous cherchons à identifier le contexte explicatif du critère consommation en carburant. Ce critère est défini par le gain de consommation en carburant réalisable sur un segment donné de l'activité de conduite. Plus le gain potentiel est important, plus le conducteur a surconsommé. Ce gain est calculé par la formule suivante :

$$Gain = 1 - (ConsoOpt/ConsoReelle) \tag{1}$$

Avec : *ConsoOpt* la consommation optimisée du segment issue d'un logiciel d'optimisation de la consommation du LTE (Laboratoire Transport Environnement de l'IFSTTAR (Mensing et al., 2013) et *ConsoReelle* la consommation réelle de carburant effectué sur le segment.

L'objectif de cette analyse étant d'expliquer ce gain en tenant compte du contexte en utilisant le processus d'assistance à la découverte de connaissances contextuelles, nous allons suivre les quatre étapes de ce processus.

5.1 Définition des composants contexte de l'activité de conduite

L'activité de conduite automobile consiste à se déplacer d'un point A à un point B au moyen d'un véhicule. Pour cela, les entités Conducteur et Véhicule interagissent ensemble dans un environnement donné qui est l'environnement de conduite (la route). Non seulement ces entités interagissent ensemble, mais elles ont une influence sur l'activité de conduite.

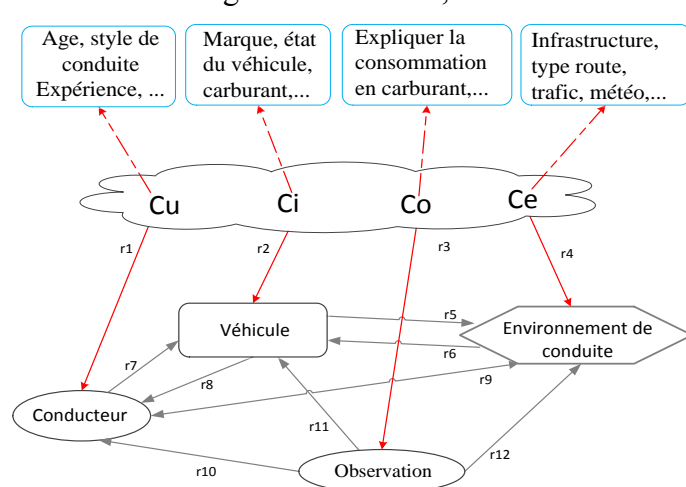


FIGURE 3 – Représentation générique du contexte de l'activité de conduite à partir de (Bazire et al., 2005).

Par exemple, en cas de trafic dense, la conduite est lente car l'environnement extérieur contraint le conducteur à tenir compte des autres véhicules. Lors de la conduite automobile, le conducteur adapte en permanence sa conduite en fonction de ce qui se passe autour de lui, donc de son environnement de conduite (trafic, météo, type de route, ...). Cette adaptation aux conditions de circulation influence les comportements ou actions de conduite. Ces changements de comportements de conduite influencent à leur tour la consommation en carburant.

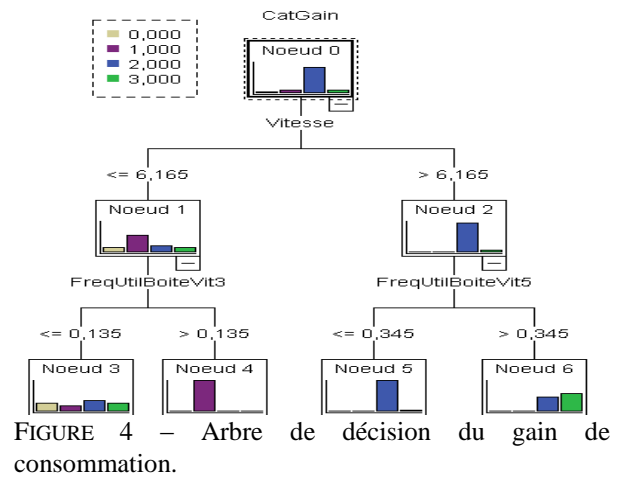
Ces informations pourraient potentiellement expliquer des différences de consommation. Elles seront donc des éléments candidats au contexte de l'activité de conduite dans le cadre de cette analyse d'après la définition du contexte de (Dey, 2001). A partir du modèle de composant contexte de (Bazire et al., 2005), nous définissons les informations candidates aux contextes des entités de l'activité de conduite automobile. Les informations relatives aux entités de l'activité de conduite (Conducteur, Véhicule, Environnement de conduite, Observateur) permettront de collecter le maximum d'informations pour mieux expliquer le critère qui est dans notre cas la consommation en carburant. Par exemple dans nos données, la conduite change en fonction du contexte de l'environnement « Ce » : type de la route (autoroute, urbain, etc.), trafic (fluide, dense, bouchon etc.), météo (neige, pluie, soleil), mais aussi du contexte conducteur « Cu » (expérience de conduite, âge, style de conduite etc.) et du contexte du véhicule « Ci » (marque, carburant, etc.). Les informations relatives à ce critère à expliquer sont définies par l'entité Observateur. Ici, on cherche à expliquer la consommation en carburant de la situation de conduite « AllerToutDroit ».

5.2 Collecte et préparation des données K/Consommation en carburant

Les données brutes collectées lors de l'activité de conduite sont des données numériques et des vidéos. Avec l'outil d'analyse de l'activité de conduite, il nous est donc possible de coder manuellement les paramètres ou variables candidates relatives au contexte de l'activité conduite grâce aux vidéos telle que : la météo, le trafic, l'infrastructure, les incidents etc. A ce niveau, il est préférable de renseigner le maximum d'informations sur les variables candidates au contexte de l'activité de la partie 5.1 et sur les autres paramètres issus des capteurs. Pour préparer les données, plusieurs opérations de préparation ont été effectuées, par exemple la création de nouveaux paramètres (calculés et contextuels). Les paramètres calculés sont définis par l'expert du domaine à partir des paramètres bruts issus de la collecte. Par exemple la moyenne de la vitesse, les fréquences d'utilisation des rapports de la boîte de vitesse sur les différents segments, le gain potentiel de consommation.

5.3 Construction de la base de connaissances K/Consommation en carburant

Cette étape a permis de définir des règles de reformulations « sûres » (requêtes basées sur des éléments objectifs, non contextuels), afin d'identifier des segments temporels homogènes en termes d'action de conduite : AllerToutDroit, TournerAGauche etc. Elle a permis également d'étiqueter ces segments de conduite en fonction de leur gain de consommation en carburant (gain calculé à partir des consommations en carburant réelle et optimale). sur l'ensemble des segments. Les segments issus de cette étape seront utilisés dans l'étape suivante pour étudier le contexte explicatif de la surconsommation dans des séquences comparables.



Nous avons utilisé une méthode de classification supervisée (Arbre de décision CRT) pour identifier les paramètres non contextuels prédisant le gain potentiel de carburant sur les segments «AllerToutDroit» de l'activité de conduite. Cette classification a permis de bien classer 90% de l'ensemble des segments utilisés pour l'apprentissage. Ce type de méthode permet d'identifier les paramètres pertinents pour séparer des segments. La méthode d'arbre de décision été choisie car elle fournit des règles de décisions mais toute autre méthode efficace pourrait être utilisée.

5.4 Processus d'assistance à l'analyse K/Consommation en carburant

Notre objectif est de montrer qu'en découvrant les variables utiles parmi les variables candidates au contexte (définies en 5.1) sur les nœuds contenant des éléments hétérogènes (10% de segments mal classés le sont dans les nœuds 3 et 6 de l'arbre de décision FIGURE 4), on arrive à améliorer le taux de classification. Nous étudions les M-Traces T_3 et T_6 issues des règles de reformulations des nœuds 3 et 6 de l'étape 5.3 avec les transformations suivantes :

$T_3 = \ll SELECT * FROM T_{sh}$
 $WHERE Vitesse \leq 6 AND FreqUtilBoiteVit3 \leq 0.14 PRODUCT CatGain = 2 \gg.$

$T_6 = \ll SELECT * FROM T_{sh}$
 $WHERE Vitesse > 6 AND FreqUtilBoiteVit5 > 0.35 PRODUCT CatGain = 3 \gg.$

Pour identifier parmi les variables contextuelles candidates, celles qui font vraiment partie du contexte, il est nécessaire à partir de l'ensemble de ces segments de créer des transformations cherchant à mieux discriminer les segments selon leur étiquette de gain. La comparaison des requêtes tentées a permis d'identifier que la variable contextuelle « *TypeInfra* » qui définit le type d'infrastructure de la route (rond-point, intersection, ligne droite, tunnel etc.) permettait de mieux classer les obsels des M-Traces T_3 et T_6 avec les règles de transformations suivantes :

$T_{Contexte1} = \ll UPDATE T_3 SET CatGain = 0$
 $WHERE TypeInfra = 'LigneDroite' OR TypeInfra = 'RondPoint' \gg.$

$T_{Contexte2} = \ll UPDATE T_6 SET CatGain = 2$
 $WHERE TypeInfra = 'Intersection' OR TypeInfra = 'Tunnel' \gg.$

Sur les obsels des deux M-Traces T_3 et T_6 , ces transformations ont permis d'en classer correctement 67% au lieu des 43% classés avec uniquement les paramètres non contextuels. En répétant ce processus sur l'ensemble des variables candidates au contexte, nous avons identifié que les variables contextuelles qui expliquent le mieux le gain de consommation sont le type d'infrastructure et le trafic.

En fin d'analyse, nous avons les résultats suivants. Avec uniquement les variables non contextuelles nous avons un taux de 90% de bonne prédiction de la classe de gain. Avec l'ajout de deux variables contextuelles ce taux est augmenté de 7% soit un taux global de bonne prédiction de 97% sur les données initiales ayant servies au processus de découverte des connaissances.

5.5 Evaluation et validation des variables contextuelles découvertes

Cette étape est importante pour savoir si les connaissances découvertes sur les données utilisées dans l'étape d'analyse des traces sont avérées sur de nouvelles données. Elle doit être effectuée avant chaque mise à jour de la base de connaissances. Pour évaluer la cohérence des connaissances produites, nous les avons utilisées sur de nouvelles données (données Test). Le taux de bonne prédiction global sur les données de test est de 84% est certes inférieur à celui de l'échantillon initial (97%), cela peut expliquer par le volume pas très important de données codées manuellement.

TABLE 1 - Test de validation des connaissances.

		Classification				
		Prévisions				
Echantillon	Observations	0	1	2	3	% correct
Initial	0	3	0	0	0	100%
	1	1	9	1	0	82%
	2	0	0	113	0	100%
	3	1	0	1	7	78%
	% global	4%	7%	85%	5%	97%
Test	0	1	0	0	0	100%
	1	0	2	2	0	50%
	2	0	2	25	0	93%
	3	0	0	2	3	60%
	% global	3%	11%	78%	8%	84%

Méthode de développement : CRT
 Variable dépendante : CatGainFuel10

Pour vérifier la stabilité des connaissances découvertes, d'autres tests de validation sont prévus sur un plus grand volume de données.

5.6 Conclusion

Le travail de recherche présenté dans cet article articule deux contributions principales dans le domaine de l'ingénierie des connaissances : **la clarification d'une démarche de**

découverte d'éléments contextuels par l'observation d'une activité tracée, débouchant sur des principes, une méthode et un environnement de découverte interactive de connaissances contextuelles ; **une démarche d'analyse originale dans le domaine de la conduite automobile** avec de premiers résultats très encourageants mettant en évidence des éléments de contexte qu'il serait difficile d'établir sans l'assistance à l'analyse mise en place. Les premiers résultats obtenus permettent d'une part, de déterminer les facteurs qui favorisent la surconsommation et constituent donc le contexte explicatif de la consommation en carburant et d'autre part, de déterminer la classe à laquelle appartient le gain de consommation potentiel sur un segment de l'activité de conduite selon les valeurs prises par les variables contexte. En appliquant le même processus sur les différents segments de l'activité de conduite, on peut déterminer les variables dont dépendent la consommation en carburant et la classe de gain à laquelle ils appartiennent. Une base de connaissances contextuelles sur la consommation en carburant est ainsi progressivement construite. Une utilisation prometteuse de cette base est la mise en place d'un système d'assistance contextuel à la conduite automobile. L'un de nos objectifs futurs est d'évaluer l'efficacité de l'approche proposée dans cet article par rapport à d'autres possibles approches de découvertes de connaissances contextuelles.

Remerciements : Nous tenons à remercier le CEESAR pour la mise à disposition des données naturelles de conduite.

Références

- AKMAN, V., & SURAV, M. (1996). Steps toward formalizing context. *AI magazine*, 17(3), 55.
- BAZIRE, M., & BRÉZILLON, P. (2005). Understanding Context Before Using It. Dans *Modeling and using context* (pp. 29-40). Berlin Heidelberg : Springer.
- BEUSEN, B., BROEKX, S., DENYS, T., BECKX, C., DEGRAEUWE, B., GIJSBERS, M., PANIS, L. I. (2009). Using on-board logging devices to study the longer-term impact of an eco-driving course. *Transportation Research Part D: Transport and Environment*, 14(7), 514-520.
- BOUQUET, P., GHIDINI, C., GIUNCHIGLIA, F., & BLANZIERO, E. (2001). *Theories and uses of context in knowledge representation and reasoning* (Technical Report No. DIT-02-010).
- BRÉZILLON, P. (2000). Modeling and Using Context: Past, Present and Future. *Decision Support through Knowledge Management*, 301-320.
- CASTELLI, G., MAMEI, M., & ZAMBONELLI, F. (2008). Engineering executable agents using multi-context systems. Dans *Engineering Environment-Mediated Multi-Agent Systems*, 5049 (pp. 223-239). Journal of Logic and Computation.
- DEY, A. K. (2001). Understanding and Using Context. *Personal and ubiquitous computing*, vol. 5, p. 4-7.
- DEY, A. K., SALBER, D., FUTAKAWA, M., & ABOWD, G. (1999). *An architecture to support context-aware applications*. GVU Technical Report GIT-GVU-99-23.
- FAYYAD, U., PIATETSKY-SHAPIO, G., & SMYTH, P. (1996a). From data mining to knowledge discovery in databases. *AI magazine*, 17(3), 37.
- FAYYAD, U., PIATETSKY-SHAPIO, G., & SMYTH, P. (1996b). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, 39(11).
- GANDON, F., & SADEH, N. (2004). Gestion de connaissances personnelles et contextuelles, et respect de la vie privée. Dans *15èmes Journées francophones d'Ingénierie des Connaissances* (pp. 5-16). Presses universitaires de Grenoble.
- GEORGEON, O. L., MILLE, A., BELLET, T., MATHERN, B., & RITTER, F. E. (2012). Supporting activity modelling from activity traces. *Expert Systems*, 29(3), 261-275.
- GIUNCHIGLIA, F., & BOUQUET, P. (1996). Introduction to contextual reasoning: an artificial intelligence perspective. Dans *Perspectives on Cognitive Science*, 3 (pp. 138-159). Sofia : B Konikof.
- GUHA, R. V. (1991). *Contexts: A Formalization and Some Applications*. Stanford University Computer Science Department.

- LAFLAQUIERE, J., PRIE, Y., & MILLE, A. (2008). Ingénierie des traces numériques d'interaction comme inscriptions de connaissances. Dans *19es Journées Francophones d'Ingénierie des Connaissances (IC 2008)* (pp. 183–195).
- LENAT, D. (1998). The dimensions of context-space. available online at URL <http://www.casbah.org/resources/cycContextSpace.shtml>.
- MATHERN, B. (2012). *Découverte interactive de connaissances à partir de traces d'activité : Synthèse d'automates pour l'analyse et la modélisation de l'activité* (thèse de doctorat en informatique). Université Claude Bernard Lyon 1.
- MCCARTHY, J. (1993). Notes on formalizing context.
- MEDDE. (2014). Observation et statistiques. Repéré à <http://www.statistiques.developpement-durable.gouv.fr/transports/r/parcs.html>
- MENSING, F., BIDEAUX, E., TRIGUI, R., & TATTEGRAIN, H. (2013). Trajectory optimization for eco-driving taking into account traffic constraints. *Transportation Research Part D: Transport and Environment*, 18, 55-61. doi :10.1016/j.trd.2012.10.003
- MILEA, V., FRASINCAR, F., & KAYMAK, U. (2008). Knowledge Engineering in a Temporal Semantic Web Context (pp. 65-74). Eighth International Conference on : IEEE.
- MULLINS, R., CARSTEN PILS, T., ROUSSAKI, D. I., & NTUA, D. (2008). Context and Knowledge Management. *Mobile Service Platforms Cluster, White paper, June*, 1–47.
- NGOC, K. A. P., LEE, Y.-K., & LEE, S.-Y. (2005). Context knowledge discovery in ubiquitous computing. Dans *On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops* (pp. 33–34). Springer.
- RAC-F. (2007, février). Changement Climatique et Transports par Réseau Action Climat-France (RAC-F). Repéré à http://www.rac-f.org/IMG/pdf/Changement_Climatique_et_Transports.pdf
- RAKOTONIRAINY, A., HAWORTH, N., SAINT-PIERRE, G., & DELHOMME, P. (2011). Research issues in Eco-driving. *Queensland University of Technology and French Institute in science and technology of transport*.
- REY, G., COUTAZ, J., & CROWLEY, J. L. (2002). The contextor: a computational model for contextual information. *Wokshop Building Bridges Interdisciplinary Context-Sensitive Computing*.
- SETTOUTI, L. S. (2011). *Systèmes à Base de Traces Modélisées : Modèles et Langages pour l'exploitation des traces d'Interactions*. (Thèse de doctorat Informatique). Université Claude Bernard Lyon 1.
- STRANG, T., LINNHOFF-POPIEN, C., & FRANK, K. (2003). CoOL: A context ontology language to enable contextual interoperability. Dans *Distributed applications and interoperable systems* (pp. 236–247). Springer.