

Apprentissage de comportements utilisateurs de produits Hypermédias

S. Bidel, L. Lemoine, F. Piat, T. Artières, P. Gallinari
LIP6, Université Paris 6
8 rue du capitaine Scott, 75015, Paris, France

Résumé

De plus en plus, les produits hypermédias (CD-Rom ou sites web) constituent des ressources riches en informations mais faiblement structurées. Pour pouvoir exploiter la richesse de tels hypermédias, il est nécessaire de proposer une aide efficace à l'utilisateur, qui lui soit adaptée. Pour cela on cherche à modéliser le comportement de navigation de l'utilisateur, qui traduit son but. Cet article traite de l'apprentissage et de la reconnaissance de comportements d'utilisateurs de produits hypermédias à l'aide de modèles markoviens.

Mots clés :

Modélisation Utilisateur, Produits Hypermédias, Modèles markoviens.

Abstract

Hypermedia such as web sites and CD-Rom are very rich but poorly structured resources where a user may easily get lost. It is then necessary to propose a user-adapted help system that allows a user to use efficiently such hypermedia. This is a problem of user modelling where we want to model the user's browsing behaviour, that is related to his goals. This paper deals with learning and recognition of user behaviours in hypermedia through the use of Markov models.

Keywords :

User modelling, hypermedia, Markov models.

1. Introduction

L'expression Anglaise 'lost in Cyberspace' a été inventée pour parler d'un phénomène auquel les utilisateurs de produits hypermédias (internet, CD-Roms) sont confrontés, vu la quantité d'informations disponibles et les moyens utilisés pour naviguer : l'utilisateur est 'perdu dans le cyberspace' quand, au cours de sa navigation, il s'éloigne de l'objet de sa recherche ou de son centre d'intérêt d'origine, jusqu'à ne plus savoir où il se trouve ni dans quelle direction chercher.

Ainsi, un des défis actuels pour les hypermédias concerne non pas la création de contenu mais l'accès efficace à ce contenu. Pour qu'un utilisateur puisse exploiter la richesse de tels produits hypermédias, il est nécessaire de lui proposer une aide. Depuis longtemps, les navigateurs incluent des mécanismes aidant à la navigation dans des structures hypermédia complexes, basés sur l'historique de la navigation (boutons *Avant* et *Arrière*) ou des moteurs de recherche « dédiés » internes au site. Ces fonctionnalités indispensables sont loin d'être suffisantes. Un système d'aide, pour être efficace, devrait pouvoir s'adapter à l'utilisateur, par exemple en ne présentant que les informations qui lui sont pertinentes ou en lui proposant des parcours plus ou moins guidés [DEB 99]. D'où le besoin de connaître ou de modéliser l'utilisateur par certains aspects [BRU 01]. Cela peut être fait à plusieurs niveaux. On peut modéliser l'utilisateur en tant que personne unique, caractérisée par son âge, ses connaissances, son métier etc. Exploitant ce type d'information, une aide « intelligente » cachera à un novice des passages trop techniques pour ne pas le noyer dans une masse d'informations, tout en permettant au contraire à un utilisateur averti d'accéder à tout le contenu mais en évitant les définitions des notions de base. Ce type d'informations sur l'utilisateur peut lui être demandé par exemple lors de sa connexion au site. A un autre niveau, on peut chercher à caractériser le comportement dynamique (i.e. de navigation) de l'utilisateur pour savoir s'il recherche une information précise ou bien s'il cherche à se documenter globalement sur un thème etc. Ces caractéristiques plus éphémères sont tout aussi précieuses et peuvent être utilisées pour lui proposer une aide pertinente. Et il n'est pas envisageable de demander à l'utilisateur de donner explicitement cette information au fur et à mesure de sa connexion (sur le Web) ou de sa session (sur CD-Rom).

Notre travail s'inscrit dans le cadre d'un projet avec un éditeur de produits multimédia. Le but est de bâtir un système d'aide pour la consultation d'hypermédias qui prenne en compte des spécificités de l'utilisateur et de ses besoins à chaque instant. Nous souhaitons concevoir de façon générique de tels systèmes, puisqu'ils sont facilement adaptables à un nouvel hypermédia. L'aide apportée à un utilisateur peut prendre plusieurs formes, Brusilovsky a par exemple proposé une liste et une taxonomie des différents types d'aide [BRU 01]. Une grande partie de ces actions d'aides sont communes à la plupart des hypermédias. On peut citer: proposer un lien de développement du thème traité par le document actuellement visualisé, proposer une interface de recherche textuelle ou une carte du site, trier les liens par ordre de pertinence, etc.

La principale difficulté dans la conception d'un système d'aide réside dans le choix de l'action d'aide à déclencher à un moment donné pour un utilisateur donné. Puisque ce choix doit être pertinent par rapport aux buts recherchés par l'utilisateur, il est nécessaire d'identifier ces buts, nous parlerons plutôt de comportements. Dans la littérature, on retrouve souvent une classification des comportements d'utilisateurs en 5 ou 6 comportements typiques [CAN 85, CAT 95, COV 88, MAR 95]. Si les termes utilisés pour ces classifications ne sont pas toujours identiques, celles-ci sont assez cohérentes entre elles. L'intérêt d'une telle taxonomie de « comportements de haut niveau » réside dans le fait que les actions d'aide à déclencher s'expriment relativement facilement à partir de ces comportements. L'identification de ces

comportements, explicités dans la section suivante, est donc la première étape logiquement nécessaire à la mise au point d'un système d'aide sophistiqué.

Le reste de cet article est organisé de la façon suivante. Après une définition des comportements, nous discutons deux stratégies possibles pour l'apprentissage, *supervisée* ou *non supervisée*. Puis nous présentons la base de données que nous avons créée et qui nous a permis de mener des expériences. Nous expliciterons également les caractéristiques locales extraites des traces de navigation et sur lesquelles travaillent nos modèles. Ensuite, nous détaillons les modèles de comportement que nous avons développés et testés. Enfin, nous évaluons les performances des différents modèles avant d'évoquer les perspectives.

2. Comportements de haut niveau et comportements découverts

Un certain nombre de chercheurs se sont penchés sur la notion de comportements de navigation dans des hypermédias, la taxonomie de [CAN 85] étant une des plus utilisées. Nous nous sommes intéressés dans notre étude à 4 des 5 comportements identifiés dans [CAN 85]. Nous avons fusionné deux comportements car ils sont proches conceptuellement et parce qu'ils devraient engendrer les mêmes actions d'aide. Nous listons ci-dessous les 4 types de comportements auxquels nous nous intéresserons, et que nous nommerons dans la suite *comportements de haut niveau* car ils sont relatifs à une vision cognitive du comportement de l'utilisateur :

- Le *Survol* consiste en la visualisation d'une proportion importante des pages constituant un thème mais sans s'y attarder. L'utilisateur cherche à acquérir une vision globale « panoramique » d'un thème.
- La *Lecture Approfondie* correspond à la lecture complète ou presque des pages visualisées, généralement dans un même thème. L'utilisateur veut se documenter de façon assez précise et complète sur un certain thème.
- La *Recherche* consiste à rechercher un document ou une information précise.
- Le *Papillonnage* correspond à un parcours désordonné sans stratégie ou but particuliers.

Disposer d'un système permettant de reconnaître ce type de comportements à partir des traces de navigation d'un utilisateur serait tout à fait satisfaisant puisque l'action d'aide à déclencher s'exprime assez naturellement en fonction du *comportement de haut niveau* supposé de l'utilisateur. On peut donc considérer que pour concevoir le système d'aide d'un hypermédia, le meilleur moyen est d'apprendre des modèles de comportement de haut niveau en mode supervisé à l'aide d'une base de données étiquetée par ce type de comportements. Mais une telle stratégie nécessite une phase de collecte et d'étiquetage d'une base de données d'apprentissage de traces utilisateur, ce qui est en général assez long et coûteux. Elle induit également une reconstruction de la base de données et un ré-apprentissage du système en cas de modification importante de l'hypermédia.

Afin d'éviter ces étapes coûteuses, nous avons choisi d'explorer le développement de techniques génériques pour découvrir et reconnaître automatiquement les comportements des utilisateurs d'un hypermédia « quelconque ». Pour cela, nous avons procédé en deux étapes.

Dans un premier temps, nous avons étudié la faisabilité de l'apprentissage de modèles pour les comportements « haut niveau », en réalisant des expériences contrôlées en mode *supervisé*, ce qui a nécessité une phase de construction d'une base de données de traces étiquetées. Ceci nous a permis d'évaluer la possibilité de modéliser et de reconnaître ces comportements, et de détecter leur changement en ligne, c'est-à-dire au fur et à mesure que la session s'écoule.

Dans un second temps, nous avons mené des expériences en mode *non supervisé*, visant à simultanément découvrir des comportements et apprendre des modèles de ces comportements. Cette approche ne nécessite pas d'étiquetage des données mais pose un problème d'interprétation des comportements découverts. Bien sûr, on espère, dans une certaine mesure, retrouver automatiquement des comportements de haut niveau tels que ceux de Canter [CAN 85] puisque cela permettrait leur interprétation (et donc le choix d'une action d'aide). Les résultats de ces expériences en mode non supervisé sont donc évalués par la correspondance entre les comportements haut niveau et les comportements découverts automatiquement.

3. Contexte d'expérimentation

Pour le développement et la mise au point de nos techniques, nous nous sommes appuyés sur l'application « L'encyclopédie du XX^e siècle », initialement un CD-Rom culturel distribué par la société Montparnasse Multimédia, et reconfiguré en site internet. Cette encyclopédie se compose d'environ 2000 articles (i.e. un article est une page contenant l'information sous forme de textes, photos, vidéos etc), d'un moteur de recherche plein texte et d'un sommaire où l'utilisateur peut naviguer dans une hiérarchie de thèmes (ou concepts) à 2 niveaux. A chaque concept est associé un ensemble de mots-clés. A chaque article sont associés un concept, des liens de navigation vers d'autres articles, ainsi qu'un temps de lecture de référence correspondant au temps nécessaire pour lire l'article dans sa totalité. La catégorisation des articles en concepts a été effectuée par les concepteurs.

3.1. Constitution d'une base de données étiquetée de comportements

L'évaluation d'un système d'aide étant extrêmement complexe et demandant des campagnes de tests, nous avons dans un premier temps réalisé une expérience contrôlée pour recueillir notre propre base de données de traces. Nous avons demandé à 16 personnes de répondre à un questionnaire d'une durée moyenne de 30 minutes et composé de 4 questions dont les réponses se trouvent dans l'encyclopédie. Nous avons réalisé deux jeux de questions différents afin d'éviter que les données collectées ne dépendent trop des particularités des questions. Le but d'un questionnaire est d'induire pour chaque question un comportement typique de façon à disposer d'une base de données permettant de réaliser un apprentissage supervisé de modèles de comportements. Les traces obtenues pour répondre à une question donnée sont donc étiquetées par le comportement de haut niveau supposé induit par la question : *Survол*, *Lecture Approfondie*, *Recherche* ou *Papillonnage*.

Par exemple, une question demande à l'utilisateur d'extraire des dates importantes d'un thème donné. Il s'agit donc de parcourir rapidement plusieurs pages d'un thème donné sans avoir besoin de les lire complètement, ce qui correspond à un comportement de « Survol ». Une autre question demande à l'utilisateur de rechercher une photo bien précise (par exemple la photo d'une explosion nucléaire). L'utilisateur adoptera donc théoriquement un comportement de type « Recherche ». Mais dans les faits, tout n'est pas aussi simple et même dans ce contexte expérimental contrôlé, quelques sujets ont avoué s'être laissés détourner de leur objectif initial par l'intérêt immédiat qu'ils portaient à un document, insérant notamment des périodes de Lecture Approfondie au milieu d'autres comportements. Ceci reflète parfaitement la difficulté de la tâche de classification du comportement, les utilisateurs changeant fréquemment de stratégie de navigation, même au cours de la poursuite d'un but unique et bien défini. Si le bruit ainsi ajouté à nos données ne facilite pas la tâche du système d'apprentissage, il leur ajoute une validité écologique, ce qui pourrait conduire à un système plus robuste dans les conditions réelles d'utilisation .

Nous avons donc récolté ainsi 64 sessions de données, 16 (une par sujet) pour chacun des 4 *comportements de Canter* [CAN 85] , ce qui après extraction des caractéristiques représente environ 400 trames représentant chacune la navigation de l'utilisateur sur une petite période (de l'ordre d'une minute) de session.

3.2. *Extraction de caractéristiques*

Une trace est une séquence temporelle d'événements. Les événements sont les changements de page (on récolte l'adresse de la nouvelle page), les clics, les scrolls, les requêtes formulées sur le moteur de recherche, etc. L'extraction de caractéristiques a pour but de calculer, à partir de ces traces utilisateur brutes, des mesures pertinentes pour la tâche de classification qui nous intéresse.

Pour distinguer les comportements de l'utilisateur, Canter et al. [CAN 85] ont proposé d'utiliser des mesures caractérisant le graphe associé au parcours de navigation : le nombre de routes suivies, de « pics » (routes suivies dans les 2 sens), de cycles formés par le parcours, etc. Ces caractéristiques ne se sont pas avérées pertinentes dans notre contexte, principalement parce qu'elles nécessitent d'être calculées sur une période de navigation relativement longue. Ce point est handicapant dans la mesure où l'on souhaite déterminer en ligne et avec un temps de réaction assez court le comportement de l'utilisateur, ainsi que les changements de comportements. Or, ce comportement peut évoluer très vite. La participation de ces caractéristiques à la performance du système restant marginale, nous ne les avons pas retenues. Parmi un ensemble de caractéristiques, nous en avons retenu 9, qui sont suffisamment génériques pour être utilisées avec divers hypermédias, et qui se sont avérées assez complémentaires. Une trame de caractéristiques est générée pour chaque page consultée, mais si son temps de consultation est inférieur à 30 secondes, les pages suivantes sont prises en compte pour le calcul de la trame jusqu'à ce que leurs temps de consultations cumulés dépassent ce seuil, ce qui amène la durée moyenne des trames à un peu plus d'une minute. Les 9 caractéristiques retenues se répartissent en 3 groupes selon le type d'information qu'elles reflètent :

Un premier groupe de caractéristiques concerne le type de *lecture* de l'utilisateur. Pour calculer ces caractéristiques, on utilise le temps de lecture de référence pour les pages visitées. On calcule un taux de lecture (rapport du temps passé sur le temps de référence) de la première partie des pages lues (la première partie de la page correspond souvent à un paragraphe introductif) ainsi que le taux de lecture pour le reste de la page (la lecture de la fin de la page nécessite de scroller). Enfin, on calcule une « activité » qui correspond au nombre d'événements survenus sur les pages lues (scroll, clic souris, ...).

Un deuxième groupe de caractéristiques est lié à la répartition du temps passé sur les différentes *ressources* disponibles : les pages feuilles contenant l'information, les pages de sommaires et la page du moteur de recherche. On calcule le pourcentage de temps passé sur ces trois types de pages, la somme de ces 3 caractéristiques vaut 1.

Enfin, un troisième groupe de caractéristiques reflète la focalisation de la navigation sur un *concept* particulier ou bien au contraire une navigation dispersée sur plusieurs thèmes. On rappelle qu'à chaque page est associé un thème, défini par un certain nombre de mots-clés. On peut donc calculer une distance entre thèmes par une mesure cosinus du type de celles utilisées en recherche d'information. La première caractéristique est définie par la similarité moyenne entre les concepts associés à deux pages successives accédées par l'utilisateur. La seconde caractéristique est similaire mais fait intervenir une pondération par le temps passé sur chaque concept. Ces deux mesures combinées permettent de déterminer si l'utilisateur s'est focalisé sur un thème, s'il a abordé de nombreux thèmes, ou s'il a changé de thème à chaque page.

4. Modèles de comportement

Nous présentons maintenant les modèles de comportement que nous avons développés et testés. Ils permettent de déterminer en ligne, i.e. au fur et à mesure de la session d'un utilisateur, le type de comportement dans lequel il se trouve. Nous nous sommes intéressés au développement de modèles statistiques basés sur des chaînes de Markov (CM) ou des modèles de Markov Cachés (MMC). Ces modèles statistiques constituent des technologies de référence dans des domaines tels que la reconnaissance de la parole, de l'écriture, de séquences biologiques etc. Ils offrent l'avantage de permettre simultanément la segmentation de séquences en différentes classes (e.g. des comportements) et la reconnaissance de ces classes. Nous avons par ailleurs testé un réseau de neurones sur nos données afin de donner des éléments de comparaison avec des méthodes de classification plus classiques.

Nous formalisons maintenant notre modélisation. La trace d'un utilisateur est transformée, par l'extraction des caractéristiques (cf. §3.2), en une séquence de trames. Chaque trame (i.e. vecteur de 9 caractéristiques) décrit localement, dans le temps, les actions de l'utilisateur. On note O une séquence de trames, et $o_1^T = (o_1, \dots, o_T)$ une séquence de T trames, o_t désignant la $t^{\text{ème}}$ trame de la séquence. Ici le temps t est utilisé comme indice des trames pour rappeler qu'il s'agit de séries temporelles. Nous nous intéressons au développement de modèles de comportements M , permettant de calculer la probabilité de traces $P(o_1^T / M)$. Nous décrivons dans la suite des modèles de comportement qui prennent en compte, ou pas, la dynamique de la

séquence. Si l'on ne prend pas en compte la dynamique, c'est à dire si l'on suppose que les trames d'une séquence sont indépendantes, on peut décomposer la probabilité d'une séquence de trames en :

$$P(o_1^T / M) = \prod_{t=1}^T P(o_t / M) \quad (1)$$

Avec cette hypothèse d'indépendance, il est possible de calculer des probabilités $P(o_1^T / M)$ à l'aide de modèles travaillant sur des trames isolées. Nous exploiterons cela en proposant un système basé sur des modèles neuronaux entraînés à discriminer entre les comportements (i.e. les classes).

On peut également relâcher l'hypothèse d'indépendance entre trames successives et chercher à exploiter la dynamique des trames en utilisant des modèles de traitement de séquences tels que des chaînes de Markov (CM) ou des modèles de Markov Cachés (MMC). Dans ce cas, la dépendance entre trames successives est prise en compte par les transitions entre les états de la CM ou du MMC.

4.1. Réseaux de neurones

Pour comparer nos systèmes dynamiques, nous avons également, en mode supervisé, développé un système basé sur un réseau de neurone de type Perceptron linéaire (Per), qui est un modèle classique pour la discrimination. Ces modèles sont appris à discriminer entre les trames des différents comportements, ils ont autant de sorties qu'il y a de comportements. Les paramètres du réseau sont appris en mode supervisé par rétro-propagation du gradient. Un réseau neuronal appris de cette façon approxime en sortie les probabilités a posteriori des classes (i.e. comportements) $P(M / o_t)$. En supposant les comportements a priori équiprobables, on a :

$$Arg \max_M P(o_1^T / M) = Arg \max_M \prod_{t=1}^T P(M / o_t) \quad (2)$$

La règle de décision consiste donc à affecter une trace au comportement maximisant le score défini dans le membre droit de l'Eq. (2).

4.2. Modèles dynamiques

Les modèles dynamiques requièrent un nombre de paramètres éventuellement important, ce qui peut poser des problèmes d'estimation. Nous avons donc, en plus des modèles standards (CMs et MMCs), développé des variantes simplifiées de tels systèmes. Ces variantes sont des *modèles multi-flux* (notés MF-CM et MF-MMC) ils sont obtenus en faisant une hypothèse d'indépendance entre les caractéristiques des différents groupes décrits en §3.2. C'est une hypothèse naturelle dans la mesure où ces groupes de caractéristiques révèlent des aspects très différents de la consultation faite par l'utilisateur. Cette hypothèse conduit à des systèmes plus simples, définis avec un nombre réduit de paramètres, et qui par conséquent devraient être plus robustes. Une trame est ainsi décomposée en trois groupes de caractéristiques, les caractéristiques liées au type de *lecture*, aux *concepts* accédés, et

aux *ressources* utilisées. Nous noterons ces trois groupes l_i , c_i , et r_i , une trame est un triplet : $o_i = (l_i, c_i, r_i)$. Avec l'hypothèse d'indépendance entre caractéristiques de différents groupes, on a :

$$P(o_i^T / M) = P(l_i^T / M) P(c_i^T / M) P(r_i^T / M) \quad (3)$$

Le modèle multi-flux d'un comportement est donc en réalité le produit de 3 modèles (CM ou MMC), un pour chaque groupe de caractéristiques, ce qui revient à considérer 3 flux parallèles de traitement de l'information. Il y a donc un modèle par groupe de caractéristiques et par comportement, et le système résultant sera noté MF- (pour Multi-Flux). Par contre, les systèmes standards (i.e. mono-flux) travaillent sur des trames entières et sont constitués d'un seul modèle par comportement.

4.2.1. Chaînes de Markov

Les CMs sont des modèles paramétriques dont il faut estimer les paramètres, c.a.d. les probabilités des états initiaux, et les probabilités de transition entre états. Le nombre d'états est choisi par essais successifs ; dans nos expériences, nous utilisons des CMs à 3 états pour chaque groupe de caractéristiques et pour chaque comportement dans le cas de modèles multi-flux et des CMs à 7 états dans le cas de systèmes mono-flux. Les CMs sont ergodiques, et de plus toutes les transitions directes entre états sont autorisées. Les états des CMs sont définis par l'estimation d'un mélange de N gaussiennes (autant que d'états) sur l'ensemble des trames, tous comportements confondus. Cette étape détermine N vecteurs de caractéristiques typiques (les vecteurs moyennes des N gaussiennes). Chaque vecteur typique est associé à un état de la chaîne de Markov. Une séquence de T trames est alors discrétisée par ce « dictionnaire » (une trame est représentée par l'état dont le vecteur de caractéristiques associé est le plus proche) et est finalement représentée par une séquence de T états. Ceci est réalisé pour chaque groupe de caractéristiques dans le cas de MF-CM. La probabilité d'une séquence de trames par une CM est calculée par la probabilité de la séquence d'états correspondante.

4.2.2. Modèles de Markov cachés

Les MMCs sont définis par une chaîne de Markov et par des lois de probabilités d'émission associées aux états. Nous avons choisi d'utiliser des lois d'émission gaussiennes, définies par un vecteur moyen et une matrice de covariance diagonale. Le nombre d'états a été choisi par essais successifs, il est égal à 3 pour les systèmes MF-, sinon à 7. Nous avons envisagé deux types de systèmes. Dans le premier type, les états (lois de probabilités d'émission) des MMCs sont partagés entre tous les modèles de comportements, nous noterons ce système MMC-P. Dans le second, tous les MMCs sont appris indépendamment. Le premier système est plus économe en nombre de paramètres et donc moins sujet à des problèmes de sur-apprentissage, mais pourrait manquer de finesse de modélisation. Ces deux types de systèmes sont utilisés dans des systèmes multi-flux et dans des systèmes mono-flux.

4.3. *Apprentissage supervisé et non supervisé*

L'apprentissage supervisé de comportements est tout à fait classique et nous avons utilisé les techniques d'apprentissage standards (Rétro-propagation, Baum-Welch etc) pour les modèles (Réseaux de neurones, Chaînes de Markov, modèles de Markov, ...) que nous avons utilisés.

L'apprentissage non supervisé consiste à apprendre des modèles de comportements sans utiliser l'étiquetage de la base de données. Cela est réalisé, pour les CMs ou les MMCs, à l'aide d'un algorithme du type EM où les données manquantes sont les probabilités d'appartenance des données (i.e. traces) aux différentes classes (i.e. comportements). On considère donc qu'une trace utilisateur peut être produite par chacun des modèles de comportements. L'algorithme alterne alors l'estimation des probabilités d'appartenance des trames d'apprentissage aux différents comportements, et la ré-estimation des modèles de comportements [CAD 00].

4.4. *Segmentation*

Dans un système d'aide, le système d'analyse du comportement est utilisé pour déterminer, au fur et à mesure du déroulement de la session, le comportement de l'utilisateur. En fonction de son comportement à un instant donné, on déclenche les actions d'aide les plus pertinentes : cacher les liens les moins intéressants, proposer des liens pertinents etc. Il est donc nécessaire de pouvoir segmenter les traces de l'utilisateur en comportements, et de reconnaître simultanément ces comportements. Ceci est réalisé simplement en construisant un modèle de Markov global incluant les modèles de comportement appris, et en définissant des probabilités de transition entre comportements. La segmentation est alors réalisée à l'aide d'un algorithme de programmation dynamique classique.

5. Résultats

Nous présentons ici des résultats expérimentaux obtenus sur notre base de données. Bien sûr, la meilleure évaluation de nos systèmes serait de les intégrer dans un système d'aide à la navigation et de collecter les avis d'utilisateurs tests. Ce type d'évaluation étant particulièrement lourd, et nécessitant une mise en œuvre complète d'un système d'aide, nous avons évalué nos systèmes de reconnaissance à l'aide de critères quantitatifs plus classiques. Nous discutons d'abord des résultats obtenus en mode supervisé, c'est à dire sur la classification de séquences en comportements de haut niveau, selon la taxonomie de Canter [CAN 85]. Puis, nous discutons des résultats en mode non supervisé. Enfin, nous donnons des résultats en segmentation.

5.1. *Mode supervisé*

Le tableau 1 compare plusieurs systèmes pour la reconnaissance des 4 comportements pour les sessions entières, d'une durée moyenne de l'ordre de 6 minutes, pour

lesquelles le comportement est supposé constant. Les systèmes sont : un système basé sur un Perceptron appris à discriminer entre les trames des 4 comportements, un système basé sur des chaînes de Markov, et deux systèmes à base de modèles de Markov cachés, utilisant des états partagés ou non. De plus, tous les modèles dynamiques sont essayés en mono-flux et multi-flux. L'ensemble de ces résultats est donné par une validation croisée sur 4 partitions des données. Les performances varient de 60 % à 80 % suivant les modèles, ce qui est relativement élevé quand on pense qu'il s'agit de reconnaître des comportements de haut niveau à partir de traces utilisateur de bas niveau. Par ailleurs, les systèmes proposés permettent aussi de reconnaître les comportements sur des séquences plus courtes. Ainsi, le système à base de MMC-P atteint 50 % de reconnaissance avec des séquences de deux trames seulement (2^e de session).

Modèle	Per	CM	MF-CM	MMC	MF-MMC	MMC-P	MF-MMC-P
% Correcte	80	66	61	75	80	66	78

Tableau 1. Taux de reconnaissance de comportements pour des sessions entières, pour 4 systèmes, un réseau de neurones (Per), des chaînes de Markov (CM) et des modèles de Markov (MMC) standards, multi flux (MF-) ou à états partagés (-P) .

On remarque la bonne performance du réseau de neurones, ce qui signifie que la valeur des trames indépendamment de leur dynamique contient déjà beaucoup d'information. En ce qui concerne les modèles dynamiques, on note la supériorité des MMCs sur les CMs, et notamment des modèles à plusieurs flux (MF-). Ceci semble confirmer que les modèles classiques, plus gourmands en paramètres, sont plus difficiles à estimer que les modèles multi-flux. Par ailleurs, le fait de partager les probabilités d'émission entre modèles de comportements (-P) peut permettre une économie substantielle du nombre de paramètres sans perte de performance dans le cas de modèles multi-flux.

5.2. *Mode non supervisé*

Comme nous l'avons déjà souligné, il serait très satisfaisant que les comportements découverts automatiquement correspondent à ceux de haut niveau. Afin d'évaluer les méthodes en mode non supervisé, nous avons mesuré la correspondance entre comportements de haut niveau (les étiquettes des données) et comportements découverts. Ces expériences ont aussi été réalisées sur des sessions entières en utilisant une validation croisée sur 4 partition. Toutes les expériences reportées ici correspondent à la découverte de 6 comportements. Ce nombre a été choisi après essais et confirmé par un calcul de la mesure F-stat sur l'ensemble des données qui prédit un partitionnement optimal avec 6 clusters. C'est un compromis entre la finesse des comportements découverts et la possibilité d'apprendre ces modèles avec notre base de données relativement limitée. Le critère de performance utilisé est assimilable à un taux de correspondance. Il est calculé comme suit. Après apprentissage, on étiquette chaque session de test par le comportement découvert dont le modèle

maximise la vraisemblance. Puis on étiquette chaque comportement découvert avec le comportement de haut niveau qui est majoritaire dans l'ensemble de sessions qui lui sont associées. On calcule alors le pourcentage de sessions pour lesquelles les étiquettes (comportement haut niveau et comportement découvert) correspondent.

Modèle	CM	MF-CM	MMC	MF-MMC	MMC-P	MF-MMC-P
% Correcte	53	61	58	61	55	59

Tableau 2. Comparaison de la correspondance pour les sessions entières entre comportements découverts et réels pour 6 modèles en mode non supervisé: Chaînes et Modèles de Markov (à états partagés -P ou non), mono ou multi flux (MF-).

On voit dans ces résultats qu'une proportion raisonnable de sessions sont bien étiquetées en terme de comportement haut niveau, entre la moitié et deux tiers. Même si on ne parvient pas ici aux 80 % obtenus en mode supervisé, les résultats reflètent les mêmes tendances et poussent aux mêmes conclusions générales, à savoir la supériorité des modèles multi-flux et la légère chute des performances en cas de partage des probabilités d'émission. Seule la performance des MF-CM surprend par rapport aux résultats précédents et se hisse au niveau de celles des MMCs, qui restent légèrement supérieures dans l'ensemble. Ces résultats sont assez prometteurs et montrent que d'une certaine façon on peut, à partir de traces utilisateur de bas niveau, apprendre automatiquement des comportements qui sont en majorité liés aux comportements de haut niveau.

5.3. Segmentation

Nous décrivons ici des expériences de segmentation d'une trace unique en comportements de haut niveau. Les modèles utilisés sont basés sur des MF-MMCs, ceux-ci ayant été les plus performants lors des tests précédents. Pour réaliser ces expériences, nous avons concaténé les traces de sessions de la base de test dans un ordre aléatoire, obtenant ainsi des sessions multi-comportements (pendant lesquelles le comportement de l'utilisateur varie). Nous avons alors segmenté ces sessions en comportements à l'aide des modèles de comportement, ce qui produit la séquence de comportements la plus vraisemblable étant donnée la trace multi-comportements. Le résultat de la segmentation est évalué par les mesures classiques de comparaison de chaînes de caractères, la distance d'édition.

Edit-distance	% Correct	% Susbt.	% Del	% Ins
MF-MMCs Supervisé	77	17	6	12
MF-MMCs Non Supervisé	41	45	14	2

Tableau 3. Comparaison des critères de segmentation pour les MF-MMCs.

Sans surprise, l'apprentissage supervisé donne toujours les meilleurs résultats, et de façon encore plus nette que précédemment. Plus de 3 fois sur 4, le système identifie correctement le changement de comportement, qu'il n'ignore que dans 6 % des cas (déletions). Le reste du temps, un changement de comportement est détecté mais n'est pas correctement identifié (substitution). L'insertion de comportements (détections de changements à tort) reste relativement limitée, à environ un cas sur 10. Ces résultats sont très encourageants car ils montrent que les changements de comportements sont assez bien détectés et que ce type de système peut permettre de piloter un système d'aide utilisateur de façon pertinente.

6. Conclusion

Dans cette étude, nous nous sommes intéressés à la modélisation du comportement de navigation d'un utilisateur d'hypermédia. Nous avons exploré l'emploi de divers modèles markoviens pour apprendre et reconnaître des comportements. Nous avons mené des expériences en mode supervisé et en mode non supervisé. Les résultats en mode supervisé ont montré la faisabilité de la reconnaissance de comportements haut niveau à partir de traces de bas niveau. Les résultats obtenus en mode non supervisé montrent que les comportements découverts automatiquement présentent une correspondance correcte avec les comportements de haut niveau. Ce résultat est prometteur dans le sens où il permet d'envisager la mise au point automatique de systèmes d'aide génériques pour des hypermédias. Enfin, les résultats de segmentation montrent que les systèmes proposés permettent de suivre en ligne le comportement de l'utilisateur au cours d'une session et qu'ils pourraient donc être utilement intégrés dans des systèmes d'aide.

Remerciements : Ce projet a été réalisé dans le cadre du projet RNTL *Gicsweb* financé par le ministère de l'industrie.

7. Références

- [BRU 01] Brusilovsky P., Adaptive Hypermédia, *User Modeling and User-Adapted Interaction* 1, 2001.
- [CAD 00] Cadez I., Gaffney S., Smyth P., A general probabilistic framework for clustering individuals and objects, *Proceedings of the Sixth ACM International Conference on Knowledge Discovery and Data Mining*, 2000.
- [CAN 85] Canter D., Rivers R., Storrs G., Characterizing User Navigation through Complex Data Structure, *Behavior and Information Technology*, vol. 4, 1985.
- [CAT 95] Catledge L., Pitkow J., Characterizing Browsing Strategies in the World Wide Web, *Computer Networks and ISDN Systems*, 1995, vol.27, No.6.
- [COV 88] Cove J., Walsh B., Online Text Retrieval via browsing, *Information Processing and Management*, 1988.
- [DEB 99] De Bra P., Brusilovsky P., Houben G-J., Adaptive Hypermedia: From Systems to Framework, *Proceedings of the ACM Computing Surveys*, Vol. 31, n° 4, December, 1999.
- [MAR 95] Marchionini G., *Information Seeking in Electronic Environments*, Cambridge University Press, 1995.