

L'ANALYSE IMPLICATIVE DANS UN CONTEXTE D'EXTRACTION DE CONNAISSANCES POUR LA MISE AU POINT DE SYSTÈMES D'AIDE À LA DÉCISION EN ANALYSE DES COMPORTEMENTS

J. PHILIPPÉ^{1,2}, T. TEUSAN^{1,2}, S. BAQUÉDANO¹, C. BOURCIER³

RÉSUMÉ

Dans cet article nous proposons une méthode basée sur l'analyse implicative permettant la mise au point d'un système d'items pour la détermination de traits comportementaux. Chaque trait est un concept que l'on va chercher à déterminer à partir d'un sous-ensemble d'items. L'intensité d'implication, la probabilité conditionnelle, le nombre d'exemples expliqués sont les indices permettant de sélectionner les items significatifs, leur poids dans la détermination des concepts et les relations d'implication entre les concepts. Nous situons cette démarche dans un cadre plus général d'Extraction de Connaissances dans les Données (ECD). Nous montrerons une application dans le cadre de la mise au point d'un système d'aide à la décision pour la détermination du besoin et de l'aptitude à la reconnaissance utilisable dans un cadre professionnel ou de formation.

SUMMARY

The implicative analysis in a knowledge data discovered for the realisation of an behaviour analysis expert system. In this paper we suggest a method based on the implicative analysis which allow to build an item system to determine the behavioural features. Each feature is a concept that we shall try to determine from an item subset. The implicative intensity, the conditional probability, the number of explained examples are the index which allow to select significant items, their weighting to determine the concepts and implicative relationship between the concepts. The process will be focused on a knowledge data discovered (KDD) operation. To realise an expert system which determine the need and the capacity at the gratitude in a professional or educational context.

MOTS CLÉS

Analyse implicative, découverte de connaissance, comportement, règles, système expert, fouille de données

KEY WORDS

Implicative analysis, knowledge data discovered, behaviour, rules, expert system, datamining

¹ PerformanSe SA Espace Performance 1 La Fleuriaye 44481 Carquefou
jacques.philippe@performanse.fr

² IRIN - équipe CID – EPN – rue Christian Pauc La Chantrerie 44306 Nantes CEDEX 3

³ ESC Rouen Bd André Siegfried 76136 Mont-Saint-Aignan CEDEX

§1. LE PROBLÈME POSÉ

Dans le cadre des principes systémiques [1](école de Palo Alto) nous chercherons à déterminer des traits de comportement qui seront ensuite utilisés à travers un système de règles pour générer un rapport qui servira ,aux spécialistes du domaine , de support pour des opérations de conseil dans le domaine de la gestion des ressources humaines

Chaque trait est assimilable à un concept. Chaque est présent chez un sujet avec une certaine importance (3 niveaux) . Pour notre application 10 traits seront utilisés :

extraversion , anxiété , combativité , affirmation , rigueur , réceptivité , dynamisme intellectuel , besoin de réalisation , besoin de pouvoir , besoin d'appartenance

Un questionnement doit permettre de les déterminer . Sa structure est la suivante :

Item_n (choix₁ , choix₂)

Au départ les experts rédigent les items et les affectent aux différents concepts de telle sorte que *choix₁* et *choix₂* fasse apparaître une opposition entre les traits ou concepts . Par exemple *choix₁* peut porter sur l'extraversion et *choix₂* sur l'anxiété .

Nous devons tenter de répondre aux questions suivantes :

- les items mesurent-ils bien ce pourquoi ils ont été posés
- quelles relations entre les concepts sont induites par le système
- quel poids à chaque item pour déterminer un concept

§2. ECD ET ANALYSE IMPLICATIVE

La méthode que nous allons utiliser pour répondre à ces questions s'appuie sur des techniques de fouille de données (datamining) et d'extraction de connaissances dans les données (ECD)

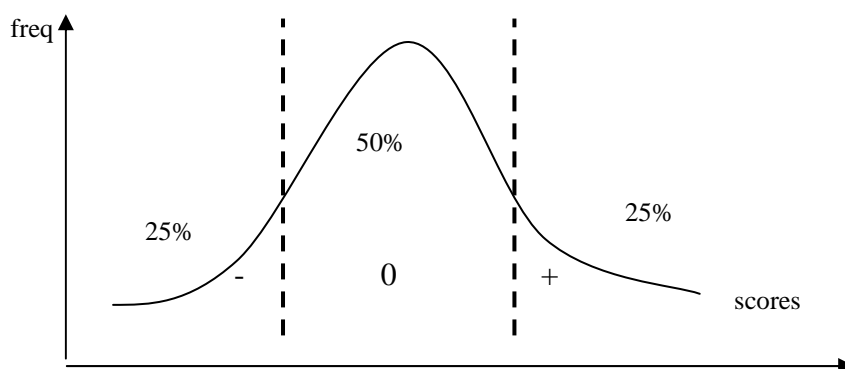


Fig. 1

Le questionnaire proposé par les experts est mis à disposition sur internet dans le but de recueillir un ensemble d'apprentissage E . Un système de scoring permet de savoir avec quelle intensité chaque sujet possède un trait suivant le principe :

Plus le sujet fait des choix dans le sens du trait plus il possède le trait

La distribution des scores sera ensuite examinée pour permettre une description de chaque trait en 3 classes suivant la représentation de la figure 1

La démarche globale que nous nous proposons de mettre en œuvre est représentée conformément à la figure 2.

La principale originalité réside dans la recherche d'implications non prévues par les experts lors de la mise au point du questionnaire. En effet les items ont été conçus pour déterminer le " trait + " par accumulation d'indices dans ce sens et cette accumulation est représentée par un score. La déduction vers le " trait - " est le résultat de " non trait + " lorsque le score est trop faible. De plus les experts ont pu observer que " trait - " et " trait + " sont présents simultanément chez un même sujet et se révèlent en fonction d'un environnement déterminé. En outre, par construction, un item peut être affecté à plusieurs traits. Il est donc tout à fait possible de découvrir des items indicateurs directs de " trait - ".

Une autre originalité consiste dans la mise en évidence des différentes façons de posséder un trait. Les experts vont chercher par exemple à découvrir les formes d'extraversion et d'introversion en répondant à la question suivante :

Quelles liens avec les autres traits impliquent les items choisis qui ont permis de déterminer l'extraversion ou l'introversion

Cette question est rendue possible en raison de la multi-affectation des items

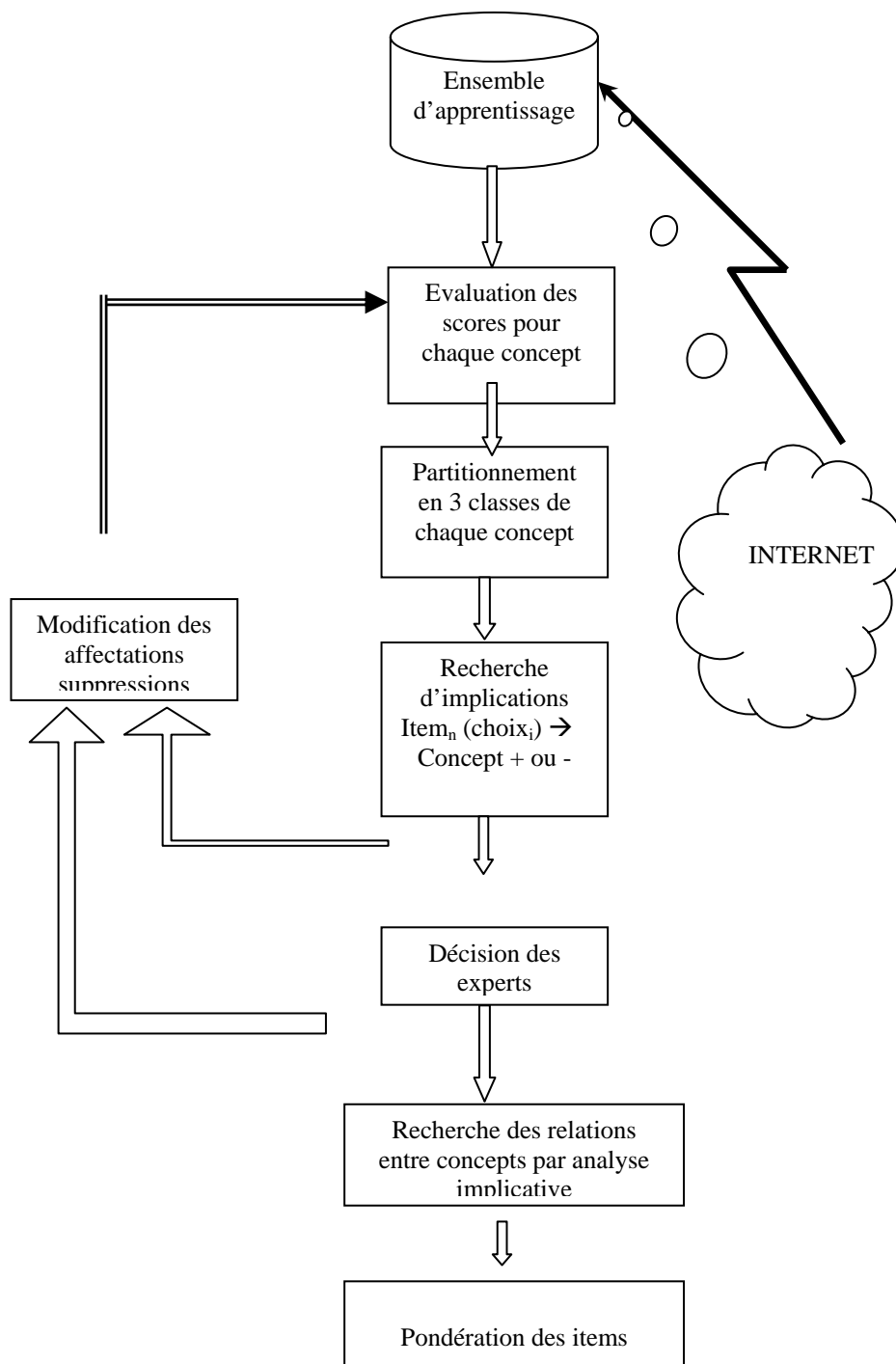


Fig. 2

§3. L'ANALYSE IMPLICATIVE

Rappelons rapidement que l'analyse implicative [11] est basée sur l'utilisation d'une mesure qui permet de déterminer la " force " avec laquelle s'établit une relation entre un descripteur (item par exemple) et une classe ou entre deux classes [3] . Cette mesure est basée sur l'étonnement statistique [9] [10] et son utilisation dans le cadre des techniques d'extraction de connaissances dans les données est très intéressante car elle permet de traiter des données bruitées . D'autre part son domaine de variation [2] et la forme de cette variation qui rappelle celle de la fonction sigmoïde utilisée pour les réseaux de neurones permet la mise en œuvre d'un système de seuillage très souple et facile à manipuler par un expert .

Nous allons donc rechercher des implications *item vers concept* caractérisés par une force d'implication f_i utilisée pour retenir les items les plus significatifs pour le concept .

Plusieurs itérations peuvent être nécessaires pour stabiliser le système d'items. Ceci réalisé une autre analyse implicative permet de caractériser les relations entre les concepts par des règles du type

$$\text{Concept}_x (-,0,+) \rightarrow \text{Concept}_y (-,0,+)$$

La découverte de ce genre de relation permet à l'expert de valider les concepts définis . L'ensemble des relations constitue un graphe : " le graphe des connaissances " dans lequel l'expert devra pouvoir naviguer .

§4. LA FOUILLE DE DONNÉES

Découvrir les relations précédentes nécessite de fouiller les données en mettant en œuvre un algorithme optimisant au maximum le parcours des sujets et des attributs qui les caractérisent. Plusieurs solutions sont possibles depuis le classique ID3 jusqu'aux algorithmes génétiques. La solution retenue est basée sur l'algorithme a-priori [6] utilisé dans l'atelier d'extraction de connaissances FIABLE-FELIX développé par la société PerformanSe à NANTES [8] . Il permet de découvrir les itemsets [7] les plus représentatifs en termes de :

- Support (nombre de sujets inclus)
- Probabilité conditionnelle d'observer un concept si un item a été choisi
- Force d'implication

Ces trois indices permettent de limiter l'explosion combinatoire liée à ce genre d'algorithme [4] [5].

§5. UNE APPLICATION

Dans le cadre d'un travail réalisé pour l'Institut de la Reconnaissance à L'ESC de Rouen nous avons utilisé cette méthode pour la mise au point d'un système d'items et de concepts devant permettre la construction d'un système d'aide à la décision pour l'analyse du *besoin de reconnaissance et de la capacité à reconnaître* dans le cadre d'une entreprise , d'un groupe , d'un cycle d'études ,

- a) Un ensemble de 128 items (2 choix) a été mis à disposition sur internet puis les résultats sont traités et soumis aux experts . Nous présentons ici un exemple de résultats sur la détermination du concept : *besoin de réalisation* (Fig. 3a, Fig. 3b) .
- Alors que les experts ont défini le concept par le + l'analyse révèle les items indicateurs du –
 - L'analyse a permis de réduire le nombre d'items (128 → 75)
 - L'analyse a permis de découvrir les multi-affectations
 - L'utilisation de la probabilité conditionnelle variant de 0.8 à 1 a permis de pondérer les items sur 3 niveaux d'importance

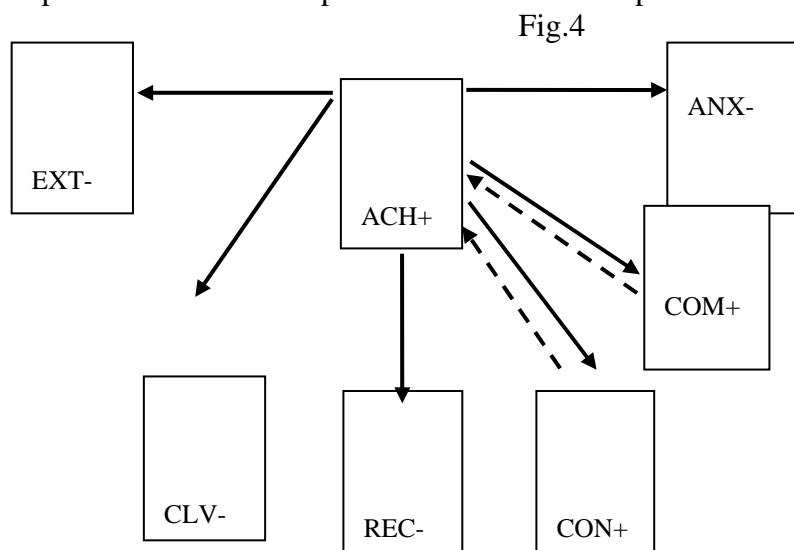
<p>Description ACH = - Seuil de probabilité conditionnelle : 0.8 Niveau maximal 2 Seuil d'intensité d'implication :0.8 Support: 27.945205479452 % Descriptions plus spécifiques</p>
<ul style="list-style-type: none"> •ITEM52 = 1, ACH = - (support = 26.02 %, proba = 93.13 %, phi = 0.91) •ITEM64 = 1, ACH = - (support = 26.02 %, proba = 93.13 %, phi = 0.99) •ITEM39 = 1, ACH = - (support = ~52.0%, proba = 90.19%, phi = 0.84) •ITEM35 = 9, ACH = - (support = 25.20 %, proba = 90.19 %, phi = 0.89) •ITEM 6 = 2, ACH = - (support = 24.93 %, proba = 89.21 %, phi = 0.99) •ITEM 125 = 2, ACH = - (support = 24.93 %, proba = 89.21 %, phi = 0.98) •ITEM 1 = 2, ACH = - (support = 24.65 %, proba = 88.23 %, phi = 0.99) •ITEM 17 = 2, ACH = - (support = 24.65 %, proba = 88.23 %, phi = 0.97) •ITEM65 = 2 ACH = - (support = 24.10 %, proba = 86.27 %, phi = 0.86) •ITEM 117 = 1, ACH = - (support = 24.10%, proba = 86.27%, phi = 0.99) •ITEM 124 = 1, ACH = - (support = 24.10 %, proba = 86.27 %, phi = 0.99) •ITEM61 = 1, ACH = - (support = 23.83 %, proba = 85.29 %, phi = 0.99) •ITEM5 = 1, ACH = - (support = 23.56 %, proba = 84.31 %, phi = 0.85) •ITEM6 = 2, ACH = - (support = 23.28 %, proba = 83.33 %, phi = 0.90) •ITEM44 = 1, ACH = - (support = 23.28 %, proba = 83.33 %, phi = 0.99) •ITEM50 = 1, ACH = - (support = 23.28 %, proba = 83.33 %, phi = 0.99) •ITEM51 = 2, ACH = - (support = 23.28 %, proba = 83.33 %, phi = 0.94) •ITEM77 = 2. ACH = - (support = 22.73 %, proba = 81.37 %, phi = 0.99) •ITEM63 = 1, ACH = - (support = 22.46 %, proba = 80.39 %, phi = 0.92)

Fig. 3a

Description ACH = + Niveau maximal 2 Seuil de probabilité conditionnelle : 0.8 Seuil d'intensité d'implication : 0.8 Support : 18.3561643835616% Descriptions plus spécifiques
<ul style="list-style-type: none"> •ITEM41 = 2, ACH = + (support = 18.35 %, proba = 100 %, phi = 0.96) •ITEM31 = 1, ACH = + (support = 18.08 %, proba = 98.50 %, phi = 0.99) •ITEM3 = 2, ACH = + (support = 17.53 %, proba = 95.52 %, phi = 0.97) •ITEM7 = 2, ACH = + (support = 17.53 %, proba = 95.52 %, phi = 0.99) •ITEM53 = 1, ACH = + (support = 16.98 %, proba = 92.53 %, phi = 0.99) •ITEM4 = 2, ACH = + (support = 16.43 %, proba = 89.55 %, phi = 0.99) •ITEM10 = 1, ACH = + (support = 16.43 %, proba = 89.55 %, phi = 0.99) •ITEM34 = 1, ACH = + (support = 16.43 %, proba = 89.55 %, phi = 0.99) •ITEM59 = 2, ACH = + (support = 16.16 %, proba = 88.05 %, phi = 0.95) •ITEM101 = 2, ACH = + (support = 16.16 %, proba = 88.05 %, phi = 0.90) •ITEM125 = 2, ACH = + (support = 16.16 %, proba = 88.05 %, phi = 0.89) •ITEM78 = 2, ACH = + (support = 15.89 %, proba = 86.56 %, phi = 0.90) •ITEM81 = 2, ACH = + (support = 15.89 %, proba = 86.56 %, phi = 0.99) •ITEM30 = 2, ACH = + (support = 15.34 %, proba = 83.58 %, phi = 0.99) •ITEM43 = 1, ACH = + (support = 15.34 %, proba = 83.58 %, phi = 0.99) •ITEM89 = 1, ACH = + (support = 15.06 %, proba = 82.08 %, phi = 0.95) •ITEM83 = 1, ACH = + (support = 14.79 %, proba = 80.59 %, phi = 0.87)

Fig. 3b

b) Nous avons ensuite effectué l'analyse des implications entre les classes pour permettre aux experts d'ajuster le concept défini par chaque classe par rapport aux autres concepts . La représentation suivante permet de se rendre compte de la complexité du réseau global



§6. CONCLUSION

La démarche implicative appliquée à ce type de questionnaire permet de mettre en évidence les relations évidentes ou cachées entre les différentes classes . La lecture des différents arbres a permis à l'expert de formuler la sémantique du domaine sous forme de textes ou de règles qu'il a pu ensuite introduire dans la base de connaissances du système expert PerformanSe-AURA. Nous concevons actuellement une démarche méthodologique complète ,s'appuyant sur ces résultats,. L'analyse des chemins menant à un trait montre qu'il existe plusieurs façons de posséder ce trait . S'appuyant sur cette constatation il devient possible d'enrichir la base de connaissances au niveau des concepts décrits . Enfin nous étudions actuellement un processus basé sur l'implication statistique pour pondérer automatiquement les items en fonction de leur participation à la détermination d'un trait.

BIBLIOGRAPHIE

- [1]- Comment une démarche de validation sur une population importante conduit à confirmer un modèle comportemental et à l'enrichir
H. Briand , S. Baquedano , L. Fleury , J. Philippé
Colloque international Techniques Psychologiques de l'évaluation des personnes, Paris (1993)
- [2]- A statistical measure of rule strength for machine learning
L. Fleury , C. Djeraba , J. Philippé , R. Gras ,H. Briand
Second World Conference on the fundamentals of Artificial Intelligence
WOCFAI 95 Paris (1995)
- [3]- Implicative Statistical Analysis
R. Gras , H. Briand , P. Peter , J. Philippé
5th Conference of the International Federation of Classification Societies ,Kobe
Japan (1996)
- [4]- Apprentissage automatique et implication : mise en oeuvre sur un espace
d'apprentissage en ressources humaines
R. Gras , F. Guillet , P. Peter , J. Philippé
4èmes journées de la société francophone de classification , Vannes (1997)
- [5]- Implicative statistical analysis
ouvrage SPRINGER VERLAG (un chapitre) (1997)
R. Gras , H. Briand , P. Peter , J. Philippé
- [6]- Contribution of the integration of implication intensity into the algorithm
proposed by Agrawal
F. Guillet , S. Guillaume , J. Philippé
EMCSR '98 14th European meeting on Cybernetics and Systems Research ,
Vienne (1998)

- [7]- L'intensité d'implication pour la découverte de règles d'association
"pertinentes"
S. Guillaume , F. Guillet , J. Philippé
RSI'98 1ères journées Ré-ingénierie des systèmes d'information 1-2 Avril , Lyon
(1998)
- [8]- Extraction de Connaissances dans une Base de Données pour la
Gestion de Ressources Humaines
FLEURY Laurent
Thèse de Doctorat de l'Université de Nantes - Novembre 96
- [9]- L'implication statistique, une nouvelle méthode d'analyse de données
Gras R., Larher A.
Mathématiques, Informatique et Sciences Humaines n°120, 1993.
- [10]- L'implication statistique, nouvelle méthode exploratoire de données.
Application à la didactique
Gras R., Ag. Almousoud S., Bailleuil M., Larher A., Polo M., Ratsimba-rajohn
H., Totohasima A.
Travaux et Thèses, édition La pensée Sauvage, 1996.
- [11]- Implication statistique et applications à l'analyse de démarches de preuve
mathématique
Larher A. Thèse de doctorat de l'Université de Rennes I, Février 1991.