

ANALYSES EXPLORATOIRES DES ACCIDENTS ROUTIERS AVEC LA CARTE AUTO-ORGANISATRICE

KONSTA SIRVIO
Sirway Ltd., Finlande
konsta.sirvio@sirway.fi

JAAKKO HOLLMÉN
Laboratory of Computer and Information Science, Helsinki University of Technology,
Espoo, Finland
jaakko.hollmen@tkk.fi

RÉSUMÉ

Les accidents routiers sont une préoccupation majeure pour la société et les individus participants. Toutes les conséquences des accidents routières motivent la recherche pour comprendre les facteurs des causalités et motivent les efforts pour minimiser des accidents. Nous recherchons les facteurs des accidents routiers en utilisant deux bases de données recordant les données sur les accidents et conditions routières aux locations des accidents en Sud de Finlande. Nous analysons les données de façon exploratoire en utilisant une Carte Auto Organisatrice. En outre, nous classons les accidents fatals en utilisant un classificateur naïf Bayes. Les risques de fatalité et injurie sont calculées pour divers conditions des accidents. Nous discutons sur nos résultats et indiquons quelques directions pour la recherche de l'avenir.

1. INTRODUCTION

Les accidents routiers sont une préoccupation majeure pour la société et les individus participants. En plus des morts humains des blessés mentaux et physiques aussi bien que des coûts financières sont causés par les accidents. Toutes les conséquences des accidents routiers clairement motivent la recherche pour mieux comprendre les divers facteurs des causalités des accidents et pour minimiser leur chiffre.

Des accidents arrivent à cause de plusieurs raisons comme erreur humain, mauvais temps et conditions routières mauvaise. Néanmoins, les causes tout explicatif ne sont pas trouvés et vraisemblablement ils ne se trouvèrent pas mais une recherche continuelle peut fournir de quelques réponses et ainsi des moyens pour minimiser les accidents.

En cette recherche relativement nouvelles méthodes d'exploration de données de l'intelligence artificielle sont présentées dans le domaine de la classification des accidents routiers et analyses des risques en utilisant les données en des bases de données publiques de la Direction Générale des Routes de Finlande. Les méthodes utilisées sont l'inférence bayésienne avec des présomptions fortes de la indépendance des données aussi bien que la Carte Auto Organisatrice (SOM) que appartient aux méthodes des réseaux neuraux.

Premièrement un réseau neural Carte Auto Organisatrice (SOM)¹ est appliqué en l'analyse exploratoire des ensembles des données. La SOM révèle la structure de

¹ Kohonen, T. (1990), pp. 1464-1480.

clusterisation des données et elle est un outil pour visualiser la structure des clusters. Deuxièmement, un classificateur naïf de Bayes est formé pour classer les accidents fatals.

2. DESCRIPTION DES DONNÉES

La Direction Générale des Routes de Finlande collectionne des données liées aux routes gardées en bases des données. En cette recherche deux sont exploités c'est-à-dire la base de données de la condition routière et la base de données des accidents routiers.²

Information about road condition measured by various variables is collected in Finland for the whole network. Collection interval depends on the type of condition variables. Most important factors of road condition are collected on yearly bases at least for the most important paved roads. These condition factors are International Roughness Index (IRI) and rutting. Average collection period for surface deflection information such as cracking and edge break is three years. Information about structural condition (central deflection) is also collected, but not utilised in this research since it is assumed that they do not have much relationship with road accidents. Table 1 summarises the variables that are taken into consideration from the road condition database.

L'information sur la condition routière mesurée par divers variables est collectée en Finlande du réseau tout entier. L'espace de la collection dépend du type des variables de la condition. Les facteurs les plus importants sont collectés au rythme annuel au moins des routes revêtues les plus importantes. Ces facteurs de condition sont l'Indice de Rugosité International (IRI) et formation d'ornières. La période de collection moyenne des défauts de la surface tel que fissuration et dégâts du bord est trois ans. L'information sur la condition structurelle (déflexion) est également collectée mais elle n'est pas utilisée en cette recherche puisque il est assumé que n'a pas un lien fort avec les accidents. En Tableau 1 présente les variables que sont pris en considération de la base de données des accidents.

Tableau 1. Variables de la base de donnée de la condition routière

1 - Route	8 - Jours du début de l'année	15 - Fissuration étroite verticale
2 - Tronçon	9 - IRI	16 - Fissuration large verticale
3 - Distance de début du début de tronçon (m)	10 - Formation d'ornières	17 - Fissuration étroite horizontal
4 - Distance de fin du début de tronçon (m)	11 - Formation d'ornières maximum	18 - Fissuration large horizontal
5 - Année	12 - Fissuration de réseaux	19 - Nids-de-poule
6 - Mois	13 - Fissuration de lignes étroites	20 - Effilochage
7 - Jour	14 - Fissuration de lignes larges	21 - Dégâts du bord

Road accident database is compiled from the information the police collects at the scene of road accidents. The database covers descriptive variables about the accident, participant information and general conditions. The variables utilised in this study from the accident database are presented in Table 2.

² www.tiehallinto.fi

Tableau 2. Variables de la base de données des accidents routiers

1 - Route	10 - Morts	19 - Luminosité
2 - Tronçon	11 - Blessés	20 - Conditions climatiques
3 - Distance (m)	12 - Type d'accident	21 - Emplacement
4 - Année	13 - Classe d'accident	22 - Participation d'un véhicule lourd
5 - Mois	14 - Chiffre de participants	23 - Alcool engagé
6 - Jour	15 - Limitation de vitesse	24 - Débit Journalier Moyen Annuel
7 - Jours du début de l'année	16 - Type de chaussée	25 - Débit Journalier Moyen Annuel de trafic lourd
8 - Jour hebdomadaire	17 - Condition de la surface	26 - Nombre des voies carrossables
9 - Heure d'accident	18 - Température	27 - Largeur du chemin

En cette recherche il était sélectionné tous les accidents routiers que s'ont produits en Finlande du Sud en territoire de Uusimaa pendant les années 1997-2005. Le tableau des accidents sélectionnés était la base en celle que l'information de condition routière était ajoutée d'après l'emplacement de l'accident. L'emplacement était déterminé par le nombre de route et de tronçon de tous les deux bases de données. Alors que les inspections de conditions étaient gérées afin que le chiffre moyen soit mesuré pour tous les segments routiers de longueur de 100 mètres, ces chiffres étaient ajoutés pour chaque accident qui avait l'information sur la distance des débuts des tronçons. Les données de condition n'étaient pas toujours cohérentes et plusieurs valeurs se manquaient. Quand ne se trouvait l'information la plus importante (IRI, formation d'ornières) pour l'emplacement d'un accident, il était éliminé de l'analyse. L'élimination était choisie dans la situation de ne s'être trouvé les valeurs de condition à l'intérieur de 200 mètres de l'emplacement et plus des 2 ans étaient passés de la dernière inspection de condition ou la prochaine inspection était accomplie plus des 2 ans du temps de l'accident.

Après tout le prétraitement de l'ensemble de données était réduite à 21 164 accidents sur que 301 accidents causaient des victimes humaines. Dans 5 058 accidents de la recherche blessures étaient rapportées. Tableau 3 résume comment l'ensemble de données était formé pour la version finale.

Tableau 3. Le chiffre d'accidents, accidents avec blessés et accidents fatals

Chiffre original d'accidents	Chiffre original d'accidents fatals	Chiffre original d'accidents avec blessés
22 935	318	5 487
Chiffre final d'accidents	Chiffre final d'accidents fatals	Chiffre final d'accidents avec blessés
21 164	301	5 058

Comme un résumé des variables ils peuvent être classifiés dans les types suivantes:

1. Variables de condition routière
2. Variables de condition du temps
3. Variables de participants
4. Variables descriptifs des accidents

3. MÉTHODES DE RECHERCHE

Prétraitement

La recherche était exécutée en diverses étapes. Dans la première étape les données originales étaient trouvées. Les étaient reçues en fichiers de texte et d'Excel. Les fichiers étaient mis dans une base de donnée dans tableaux séparés. Une interface était programmée en C++. Le processus de sélection des valeurs de condition routières corrects pour chaque accident était accompli avec langage structuré de requêtes (SQL) des bases de données combiné avec boucles plus compliquées programmées avec C++.

Après avoir terminé les tableaux de données ils étaient exportés de la base de données dans fichiers de texte et depuis importés à Matlab –programme que était utilisé en la analyse finale. Quelques valeurs absents faisant notés comme température mesurée, catégorie de luminosité, conditions de temps et trafic moyen. Les variables ceux-là aussi bien que les accidents avec les valeurs absentes étaient inclus. Les valeurs absentes étaient remplacées par zéro. Toutes les données se faisant normalisée par Équation 1. Après tous les variables avaient zéro comme la valeur moyenne et un comme déviation standard.

$$Z = \frac{X - \mu_v}{\sigma} \quad (1)$$

Carte Auto Organisatrice

Carte Auto Organisatrice est un réseau neural artificiel³ formé en manière non surveillée c'est-à-dire sans labels de classe. La Carte révèle la structure naturelle dans les données et visualise la structure des clusters de données de la haute dimensionnalité en écran de base dimensionnalité.

Les procédures de formation de la Carte Auto Organisatrice sont divisées en deux étapes que sont appliquées en manière alternante : d'abord un échantillon est extrait de la base de données et le neurone (cellule de carte) le plus proche dans SOM est cherchée comme le vecteur le plus près dans l'espace d'entrée. Depuis, le neurone vainqueur et les neurones dans le voisinage topologique sont mis à jour pour être plus proche à l'échantillon. Ce processus est répété millier fois et la carte converge à représenter les données d'entrée (avec paramètres de formation convenablement choisis).⁴

La méthode rencontre la structure de données interne sans aide d'un enseignant externe contrairement à plusieurs autres réseaux neuraux. Comme la résultat, la carte avec le nombre des neurones sélectionné est produite. La carte peut être utilisée pour visualisation aussi bien que pour trouver des clusters possibles dans les données en façon visuelle ou informatique. Les clusters naturels peuvent être examinés par soi-disant matrice U que révèle les distances entre neurones individuels.⁵

Agglomération de moyens K

La idée de agglomération est à former clusters similaires des données d'entrée. La similarité peut être mesurée par deux critères basiques – la distance minimum entre le point du centre du cluster et les autres points dans le cluster aussi bien que la distance maximum entre des clusters séparés. Le ratio de ces variables nommé la variation à

³ Kohonen, T. (1990), pp. 1464-1480.

⁴ Haykin, S. (1999), pp. 446-466.

⁵ Ultsch, A. and Siemon, H. (1990), pp. 305-308.

l'intérieur du cluster $wc(C)$ et la variation entre clusters $bc(C)$ peuvent être utilisées comme la fonction score pour mesurer la qualité de agglomération.

L'agglomération de moyens K est basiquement équivalent à l'algorithme de SOM avec l'exception q'une fonction voisinage n'est pas utilisée hormis seulement le neurone gagnant est mis à jour. En cet algorithme un nombre de clusters K sont soit sélectionné soit choisi au hasard. Les centres des clusters sont les valeurs moyennes des vecteurs de données appartenant à un cluster. Chaque point de donnée est assigné à un cluster où la distance Euclidienne entre le centre et le point de donnée est minimisé. Les centres des clusters sont mis à jour après chaque tâche d'un point de donnée et l'algorithme est répété jusqu'à ce que il n'y aura plus des changements, par conséquent résultant au moins à un optimum local.⁶

Classification naïve de Bayes des accidents fatals

En classification naïve de Bayes les labels des classes sont connues et la formation du classificateur naïf de Bayes est accompli par le choix des variables et l'estimation des distributions des classes. En notre exemple les deux groupes des classes suivants sont analysés :

1. Accidents fatals et accidents non fatals
2. Accidents avec blessés et accidents sans blessés

En classification naïve de Bayes les variables observés sont assumés indépendant conditionnellement donnée l'information sur classe. Bien que les suppositions soient souvent transgressées en pratique, la classification fonctionne bien malgré le défaut.

En inférence de Bayes un variable aléatoire peut être classé en catégories selon l'information précédente sur les résultats et valeurs mesurées des variables dépendants du phénomène. Dans le modèle plein de Bayes toutes les combinaisons des fonctions de distribution de probabilité jointe entre les variables sont modelées et pries en considération en calculs. Étant considérablement laborieux avec problèmes larges uniquement quelque unes peuvent être modelées et en autre extrême il y a la méthode nommée Bayes Naïf qui tient des suppositions fortes sur les variables puisque ils sont présumés indépendants entre l'un l'autre. Donc, il ne faudra pas calculer les fonctions de distribution de probabilité jointe.

Annoncemos une occurrence du phénomène comme un accident fatal par FA et non fatal par NFA. En ce cas, la probabilité d'un accident fatal étant donnée les données de mesurage X est présentée en Équation 2.⁷

$$P(FA | X) = \frac{P(FA)P(X | FA)}{P(X)} \quad (2)$$

Ici $P(X) = P(X | FA)P(FA) + P(X | NFA)P(NFA)$ que résulte à Équation 3.

$$P(FA | X) = \frac{P(FA)P(X | FA)}{P(X | FA)P(FA) + P(X | NFA)P(NFA)} \quad (3)$$

⁶ Hand, D; Mannila, H; Smyth, P. (2001), pp. 296-308.
⁷ Hollmén, J. (2000), pp. 16-18.

L'Équation 3 peut alors être utilisée en calculs du risque de fatalité ou blessure tenant compte tous les vecteurs de variable aléatoire X mesurés ou collectionnés présumptueux leur indépendance.

4. RÉSULTATS

Toutes les expérimentations étaient réalisées utilisant SOM ment du logiciel Matlab.⁸

Premièrement, la combinaison des données de accident et condition était traité par la Carte Auto Organisatrice et une carte de distance (matrice U) entre les neurones était fabriquée. Possible bordures de cluster étaient révélés et une question était soulevée si les accidents que avaient causé fatalités et blessures se différenciaient des autres accidents d'une certaine façon. En placeant sur la carte les accidents avec blessures (points noirs) ou fatalités (points rouges) il peut être vu que les accidents fatals étaient groupés proche l'un l'autre pendant que les accidents avec blessures fussent largement séparés comme illustré à gauche de Figure 1.

⁸ Vesanto, Juha and Himberg, Johan and Alhoniemi, Esa and Parhankangas, Juha. (2000)

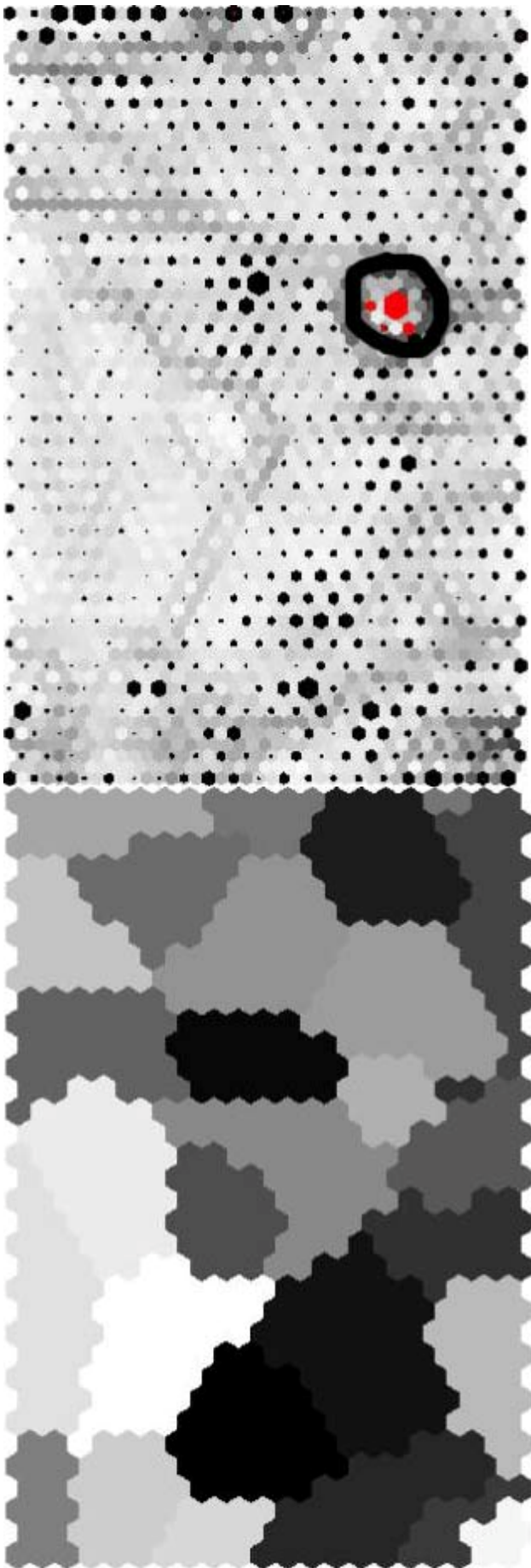


Figure 1. Matrice U de SOM avec les accidents fatals marqués avec les points rouges à l'intérieur du cerclé à gauche et bordures de clusters à droite

Les couleurs plus foncées dans la figure à gauche indiquent possible bordures de clusters. Comme les accidents fatals étaient groupés proche l'un l'autre, certaine clusterisation peut être attendue. En conséquence, le lot entier des accidents s'était rassemblé afin de voir si les accidents fatals formaient un groupe uniforme. La clusterisation de moyens K était appliquée à la carte de neurones générée auparavant sans restriction du nombre de clusters. La base pour l'agglomération était les distances entre les neurones (matrice U). L'algorithme de clusterisation a produit 27 clusters de lesquelles une clairement contenait la plupart des accidents fatals. Ce cluster comprenait 294 accidents et ainsi 7 accidents fatals étaient inclus dans un autre cluster.

Les valeurs de variables normalisées pour le cluster des accidents fatals étaient examinées par une représentation non symétrique des données en dessinant les valeurs minimums et les valeurs de 95% centile des valeurs de variable. Les valeurs normalisées étaient dessinées avec les mêmes indicateurs pour l'ensemble de données entier et les résultats sont illustrés en Figure 2. Ici, la ligne solide symbolise les données du cluster et la ligne avec plus-symboles représente l'ensemble de données entier.

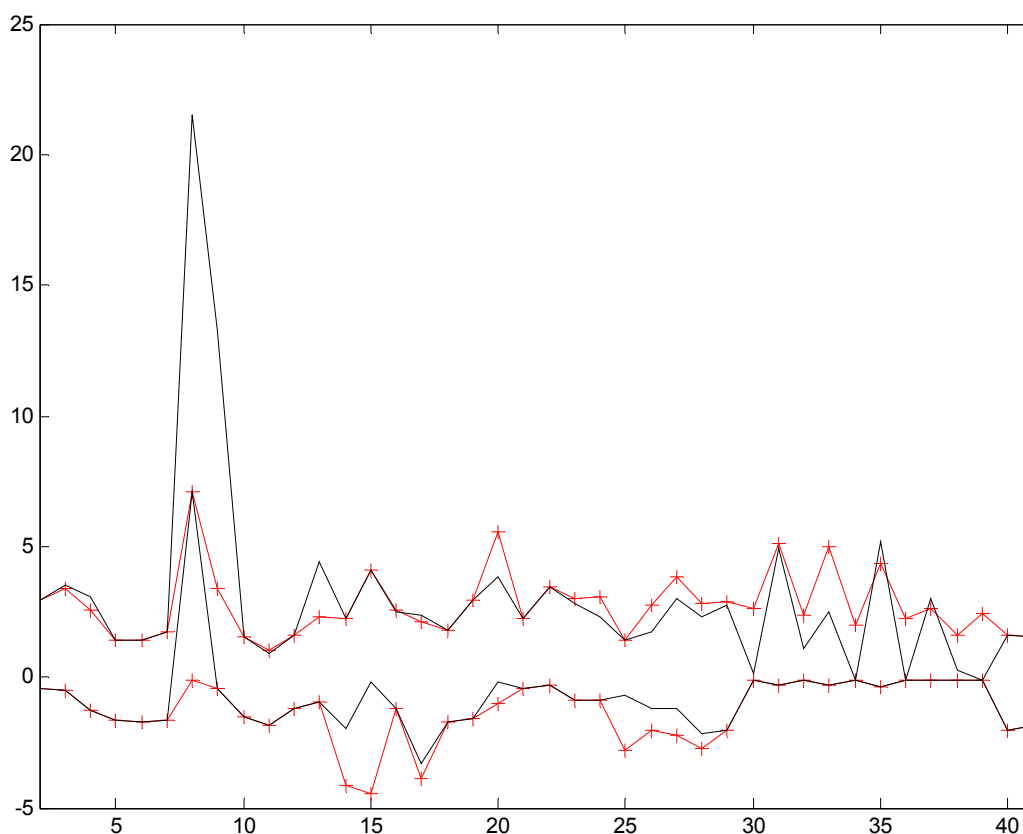


Figure 2. Valeurs minimums et 95% centiles pour les données de cluster et pour toutes les données normalisées

Les probabilités étaient calculées pour chaque variable afin de mieux obtenir la idée des caractéristiques des accidents fatals. En premier, les probabilités de classe étaient calculées pour les cas : Les accident fatals et non fatals si un accident est produit. Depuis, les probabilités des résultats variés dans cetttes classes étaient calculées. Ensuite, un classificateur naïf de Bayes était appliqué dans le but de découvrir quels types de accidents sont fatals avec une probabilité plus haute et vice versa.

Puisque la théorie de Bayes Naïf assume une hypothèse forte d'avoir les variables indépendantes quelques variables étaient exclues avoir eu trop de corrélation avec quelque autre variable. Formation d'ornières et formation d'ornières maximum mesurées étaient hautement corrélées. Les défauts de surface n'étaient pas mesurés uniformément sur tout le réseau. Le type d'accident passe la même information que la classe d'accident mais en forme mais détaillée. L'année d'accident, tronçon et distance du début de tronçon n'étaient pas considérés pertinents. Nombre de jours du début de l'année tenant essentiellement la même information que le mois. Le type de chaussée n'avait pas beaucoup d'effet au risque de fatalité aussi bien que l'alcool. 7,6% de tous les accidents qui ont causé des fatalités étaient conduits sous l'influence d'alcool pendant que 7,9% de tous les accidents tînt un chauffeur avoir consommé de l'alcool. L'heure d'accident ne semblait avoir eu beaucoup d'effet à la fatalité à part d'une particularité des accidents qui arrivaient entre 15 :00 et 16 :00. Aux environs de 8,6% des accidents fatals se passaient pendant les heures celles-ci pendant que 5,8% des accidents non fatals se ment pendant les mêmes heures.

Tableau 4 résume les caractéristiques et conditions des accidents où les probabilités de classe favorisent les accidents fatals.

Tableau 4. Probabilités conditionnelles des variables dans différentes catégories de accident

	P (X Accidents fatals)	P (X Accidents non fatals)	P (X Blessés)	P (X Non blessés)
P (catégorie accident produit)	1.42 %	98.58 %	23.90 %	76.10 %
Nombre de la route = 1,2,4,7,25,51,130,148	51.85 %	34.48 %	30.38 %	36.09 %
Chiffre des voies carrossables = 1	72.76 %	67.00 %	64.91 %	67.70 %
80 cm < Largeur de la route < 160 cm	57.15 %	47.19 %	43.91 %	48.41 %
Limitation de vitesse => 80	79.40 %	64.85 %	55.24 %	68.14 %
0.75 < IRI < 1.5	67.78 %	56.47 %	53.50 %	57.61 %
Formation d'ornières < 8	55.48 %	51.74 %	50.53 %	52.19 %
500 < Trafic lourd < 1700	51.83 %	37.43 %	36.30 %	38.05 %
4000 < Trafic < 17000	52.49 %	36.60 %	35.84 %	37.14 %
Mois = mars, avril ou juin-septembre	53.48 %	44.19 %	50.26 %	42.46 %
Jour hebdomadaire = jeudi ou vendredi	33.22 %	30.98 %	30.72 %	31.10 %
11 < Jour < 17	20.60 %	16.58 %	16.13 %	16.80 %
Temps = clair	40.86 %	35.69 %	40.43 %	34.30 %
Luminosité = lumière du jour	62.79 %	50.52 %	63.86 %	46.56 %
16 < Température < 36	26.58 %	17.97 %	24.44 %	16.10 %
Surface = seche et nette	58.14 %	52.50 %	54.84 %	51.88 %
Classe d'accident = rencontre des véhicules	35.22 %	3.33 %	8.54 %	2.29 %
Participant lourd engagé	38.21 %	15.91 %	11.39 %	16.59 %
Participants > 2	13.63 %	7.74 %	100.00 %	6.71 %
Blessés > 0	45.83 %	23.55 %	15.09 %	0.00 %

La première rangée informe que si un accident se produit la probabilité pour avoir blessés est 1,4%. Le premier groupe de variables caractérise la route où les accidents se passent. Dans le seconde catégorie l'information du temps d'accident est donnée. Le troisième groupe fournit l'information sur les conditions de temps. Le dernier groupe caractérise les accidents soi-même.

Quelques points intéressants se peuvent être notés. La plupart d'accidents fatals en Finlande du Sud se produisent en 8 routes. La fatalité était augmentée en routes étroites où la limitation de vitesse était la plus haute. Le risque de fatalité s'augmentait aussi en chemins avec les valeurs d'IRI et formation d'ornières inférieures c'est-à-dire en chemins

de bonne condition. Il se passait aussi en endroits avec Débit Journalier Moyen Annuel intermédiaire pour tout le trafic et véhicules lourds. Quelque fluctuation du risque de fatalité temporel était observée puisque des accidents fatals se produisent plus vraisemblablement au milieu de mois que les autres accidents. Également, la plupart d'accidents fatals s'ont produit en printemps, été et début d'automne.

En prévention d'accidents la priorité devrait être mis afin que les accidents dont probabilité de fatalité est plus haute que non fatalité se soient attelés avant tout. En utilisant Équation 3 et remplaçant $P(FA)$ par 0,0142, $P(NFA)$ par 0,9858 et $P(X|FA)$ aussi bien que $P(X|NFA)$ par les probabilités calculées en Tableau 1, les risques de fatalité peuvent être calculés. Quand diverses conditions sont conformes, les probabilités conditionnelles $P(X|FA)$ et $P(X|NFA)$ sont produits des probabilités des conditions individuelles selon supposition forte de l'indépendance des variables.

La division des accidents dans autres catégories en cas des accidents fatals et non fatals aussi bien que accidents de blessures et non blessures sont présentés en Tableau 5.

Tableau 5. Probabilités conditionnelles des diverses classes d'accident

Classe d'accident	Fatals	Non fatals	Blessés	Non blessés
Accident individuel	26.58 %	25.46 %	34.58 %	22.62 %
Accident de virage	3.32 %	4.67 %	7.75 %	3.68 %
Accident de dépassement	3.65 %	5.81 %	4.17 %	6.28 %
Accident de croisement	6.31 %	6.61 %	11.31 %	5.13 %
Rencontre des véhicules	35.22 %	3.33 %	8.54 %	2.29 %
Rencontre en arrière	2.33 %	10.62 %	12.69 %	9.81 %
Accident de moto	1.33 %	1.18 %	3.28 %	0.53 %
Accident de vélo	5.65 %	1.69 %	5.40 %	0.60 %
Accident de pédestre	10.3 %	1.02 %	3.54 %	0.40 %
Accident d'élan	2.66 %	10.98 %	4.63 %	12.82 %
Accident de cerf	0.00 %	23.53 %	0.67 %	30.27 %
Accident de autres animaux	0.00 %	1.12 %	0.38 %	1.33 %
Autre accident	2.66 %	3.97 %	3.06 %	4.23 %

En Tableau 6 un exemple est montré comment les risques de fatalité et blessure augmentent dans un accident ayant croissante chiffre de conditions en vigueur.

Tableau 6. Risques de fatalité et blessure sous diverses conditions

Condition	Risque de fatalité	Risque de blessure
Probabilité de catégorie	1.42 %	23.90 %
& Classe d'accident: Rencontre des véhicules	13.22 %	53.96 %
& Participant lourd engagé	26.79 %	51.59 %
& Plus des 2 participants	39.18 %	64.40 %
& Nombre de route: 1,2,4,7,25,51,130 ou 148	49.21 %	60.37 %
& IRI € [0.75, 1.5]	53.77 %	58.58 %
& Trafic € [4 000, 17 000]	62.52 %	57.43 %
& Luminosité = lumière du jour	67.46 %	64.92 %

Quand les 5 premières conditions de Tableau 6 et au moins une autre condition de Tableau 5 sont en vigueur, il est plus probable d'avoir victimes que simple blessures ou

dommages aux biens. De plus il est intéressant à noter que le risque de blessure reste plus stable mais supérieur à 50% dans la classe d'accident de rencontre des véhicules.

5. DISCUSSION

Les résultats montrent que le risque de fatalité est augmenté le plus quand l'accident routier est soit avec un piéton ou rencontre des véhicules. Il est naturel ayant généralement la vitesse relative la plus haute entre les véhicules en rencontres et dans le deuxième cas les piétons n'ont pas beaucoup de protection contre des masses plus des 1 000 kg roulant en vitesse relative plus haute. Et comment se peut réduire le risque ? Une possibilité serait ayant des rampes qui séparent les voies des directions opposées ou fabriquant voies carrossables séparées. Un nombre d'accidents relativement bas étaient classé comme accidents de dépassement et beaucoup plus étaient dans la catégorie de rencontre des véhicules. Il y a une pléthore de raisons. Ils peuvent être suicidés, sommeil, attaque, route glissante ou lecture des messages du téléphone mobil.

Comme des accidents produits au réseau de La Direction Générale des Routes de Finlande étaient examinés ils arrivaient habituellement aux emplacements ruraux. Puisque il y eût des accidents de piéton on peut assumer que les victimes étaient en train de marcher sur les côtés des routes et des voies pour trafic léger n'existaient pas. Une solution simple est à construire des voies séparées pour trafic léger à côté des routes quand trafic léger est attendu.

Les accidents avec véhicules lourds naturellement augmentent le risque de fatalité. Il n'existe pas une solution simple mais diverses méthodes se devraient être utilisées tel que favorisation du transport des biens en réseau de chemin de fer, meilleur contrôle des durées de conduite et conditions des véhicules ou voies séparées pour les directions opposées.

Quand les accidents avaient plus des 2 participants le risque de fatalité s'augmentait grandement. Également le risque de fatalité s'avait augmenté sous les conditions où l'IRI mesurée était entre 0,75 et 1,5 ; formation d'ornières sous 8 et quand le trafic était entre 4 000 et 17 000. Néanmoins, sous les mêmes conditions le risque de blessure était amoindri. Bonnes conditions routières peuvent donner une fausse impression de sécurité ainsi provoquant un excès de vitesse. L'explication est soutenue par ayant le risque de blessure plus haut en lumière de jour avec une surface sèche et nette en été quand le temps est clair et la température entre 16 et 36 degrés. Cependant, entretenant les routes en condition inférieur augmenterais les coûts de usagers routiers.

Les routes relativement étroites (80-160 cm) et l'existence de seulement une voie carrossable augmentaient le risque de fatalité. Les routes avec deux voies carrossables réduirait le risque de fatalité et rencontre des véhicules.

Selon les résultats de la recherche nous suggérons que les Directions des routes agissent en les points suivants en routes les plus hasardées quand faisable :

1. Minimisation la probabilité de rencontre des véhicules des directions opposées par voies carrossables séparées ou voies supplémentaires avec une rampe au milieu
2. Élargissement des routes hasardées
3. Diminution du trafic de camion entre trafic plus léger par une voie exclusive pour le trafic lourd, transfert du transport des biens aux chemins de fer et méthodes de contrôle plus strict sur les conditions de véhicule et chauffeur
4. Voies séparées pour le trafic léger quand dès que possible

5. Campagnes sur la sécurité de trafic en printemps et été

6. RECHERCHE FUTURE

Dû à la complexité du phénomène il se peut ajouter plus des données dans l'analyse afin de trouver plus des causes de accident et augmenter la fiabilité des probabilités de risque. Quelques données il se peut acquérir plutôt facilement. Elles sont les variables de la géométrie routière parmi lesquelles le pourcentage des montées et baisses, courbure horizontale et élévation. La géométrie routière affecte la visibilité et conduite d'un véhicule et ainsi elle peut être une cause à quelques accidents. La base de données de accidents contient plus d'information sur les participants comme âge et sexe en tableaux séparés. Avec certain travail supplémentaire ils peuvent être compris. Autres données comme mesures de temps historiques pourraient être ajoutées.

Une recherche intéressante comparative serait avec inclusion des régions de Finlande séparées et enquête s'il y a quelques différences entre les causes d'accident et dues de quelles raisons. Toutes les données du pays pourrait être groupées, une analyse similaire conduite et depuis comparée à autres pays avec les données similairement collectionnées. Il pourrait révéler les meilleures pratiques en prévention des accidents dans divers pays.

Des méthodes plus précises peuvent toujours être développées et présentées en recherche. Ici, la méthode de Bayes Naïve était utilisée. Il se pourrait que les suppositions ayant l'indépendance entre les variables soient trop fortes et ainsi un réseau de Bayes complète pourrait être utilisé quand une dépendance forte entre les variables existe si la puissance computationnelle nécessaire n'est pas une entrave.

RECONNAISSANCES

Nous voudrions remercier Ismo Iso-Heiniemi, Raija Huhtala et Markku Visti de La Direction Générale des Routes de Finlande ayant fourni les données originales. D'ailleurs, une gratitude prédominante est adressée à la Fondation de Culture de Finlande pour une bourse pour les études doctorales. Outre le groupe indiqué ci-dessus nous voudrions remercier tous qui nous ont aidé dans le processus de recherche.

RÉFÉRENCES

1. Kohonen, Teuvo. (1990). The Self-Organizing Map. Proceedings of the IEEE volume 78 number 9
2. www.tiehallinto.fi
3. Kohonen, Teuvo. (1990).
4. Haykin, S. (1999). Neural Networks. A comprehensive foundation, Prentice-Hall Inc. London, United Kingdom
5. Ultsch, A. and Siemon, H. (1990). Kohonen's Self-Organizing Maps for exploratory data analysis. Proceedings of the International Neural Network Conference (INNC'90). Kluwer
6. Hand, D and Mannila, H and Smyth, P. (2001). Principles of Data Mining. MIT Press. Cambridge, United States of America
7. Hollmén, Jaakko. (2000). User profiling and classification for fraud detection in mobile communications networks. Helsinki University of Technology. Espoo, Finland
8. Vesanto, Juha and Himberg, Johan and Alhoniemi, Esa and Parhankangas, Juha. (2000). SOM Toolbox for Matlab 5. Helsinki University of Technology, Laboratory of Computer and Information Science. Espoo, Finland (Software freely available via WWW at URL: <http://www.cis.hut.fi/projects/somtoolbox/>)