

Optimisation d'un logiciel de prévision d'avalanches à l'aide d'un algorithme génétique.

Ing.T. LENFANT
Ir R. LESCROART
PIERRARD - Virton

En Ecosse, l'évaluation des risques d'avalanche se fait à l'aide d'une méthode statistique basée sur le « Principe des plus proches voisins ». C'est un algorithme génétique qui ajuste les paramètres du modèle.

Ici, nous proposons quelques améliorations qui mènent à une meilleure exploitation des données disponibles. Les erreurs de prévision sont ainsi réduites de près de 20%.

Mots-clefs : risque d'avalanches, prédiction, algorithme génétique, ajustage de paramètres.

In Scotland, the forecasting of avalanche risk is done with the help of a statistical model, the so-called "Nearest Neighbours Method". A genetic algorithm performs the tuning of the model.

In this paper we show how to improve avalanche prediction by making a better use of the available information.

Keywords : avalanche risk, forecasting, genetic algorithm, model teening.

Introduction

Courir la montagne est une activité très populaire en Ecosse. C'est pourquoi un centre de prévision des avalanches a été mis en place, le Scottish Avalanche Information Service (SAIS). Dès l'apparition de l'informatique, des modèles de prévision se sont développés, mais à cause de la complexité du manteau neigeux, seuls les modèles statistiques se sont avérés réellement exploitables. Ces modèles se basent sur le principe : "à conditions semblables, conséquences semblables".

Dans le cadre d'un échange Socrates - Erasmus, à l'Université d'Edinburgh, il a été proposé d'étudier un modèle de prévision existant, développé spécialement pour l'Ecosse et utilisant un algorithme génétique pour sa configuration.

1. Les avalanches en Ecosse

1.1 Généralités

L'escalade en hiver concerne chaque saison environ 800 000 passionnés dont la plupart sont des citoyens. C'est à eux seuls que sont destinées les prévisions d'avalanches car les montagnes écossaises ne sont pour ainsi dire pas habitées. Cette situation est donc différente de celle des pays alpins où lorsque les risques d'avalanche sont trop importants, il peut être nécessaire d'évacuer routes et villages.

1.2 Le climat en Ecosse

Pour bien comprendre la nécessité de développer un logiciel spécifique à l'Ecosse, il est important de souligner les différences entre les conditions climatiques écossaises et suisses (prises comme représentatives des Alpes).

En Suisse, le climat est continental tandis que l'Ecosse est soumise à un climat maritime. Cela entraîne une différence de comportement du manteau neigeux et donc des conditions de déclenchement des avalanches.

Ainsi, les périodes de chutes de neige sont, en Suisse, plus longues et plus espacées qu'en Ecosse qui connaît de nombreuses petites précipitations. En

période de gel, l'Ecosse, contrairement à la Suisse, connaît d'importantes fluctuations de température. Toute proportion gardée, les avalanches sont alors plus fréquentes en Ecosse.

L'Ecosse est également balayée par des vents de grande intensité. D'une vitesse moyenne de 47 km/h, ils peuvent atteindre 246 km/h et entraîner la formation de congères bien plus importantes qu'en Suisse.

1.3 Prédiction des avalanches en Ecosse

En raison des particularités climatiques et nivologiques écossaises, le SAIS a divisé les régions montagneuses en cinq zones différentes qui, comparées à celles définies en Suisse ou en France, paraissent très petites.

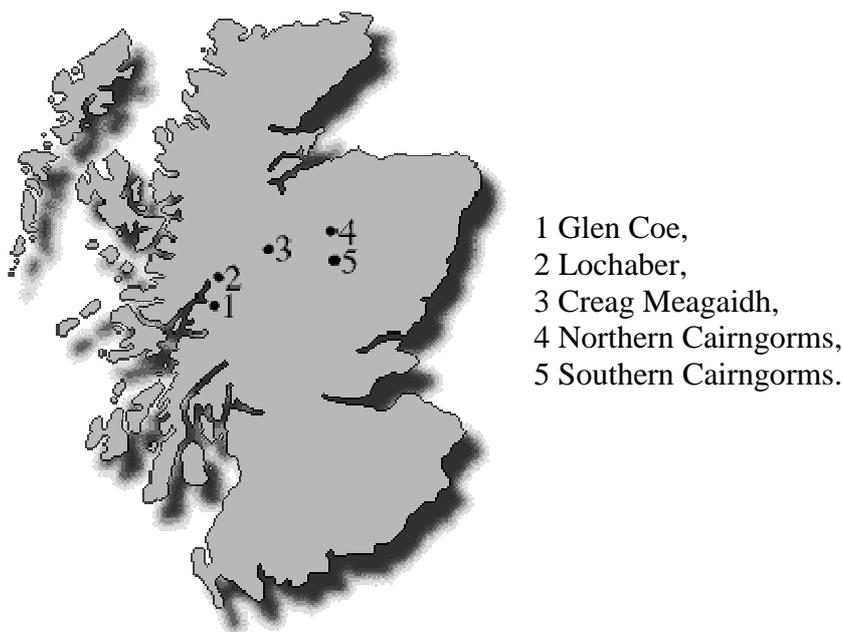


Figure 1: Zones de prévision d'avalanches.
[www.SAIS.gov.uk]

Pour chaque zone, un bulletin journalier informe les montagnards de la situation du manteau neigeux, des prévisions météo et du risque d'avalanche pour les 24h suivantes. Le bulletin est rédigé par des prévisionnistes qui, pour leur analyse, se basent sur les caractéristiques de stabilité du manteau neigeux et sur les données météorologiques concernant leur zone. Ils confrontent ensuite leurs impressions aux résultats fournis par Cornice, un mo-

dèle statistique qui fait appel à la méthode des plus proches voisins. Pour finir, ce bulletin est affiché et distribué dans les principaux points de passage de la région (auberges, office du tourisme, refuges, etc.).

2. La Méthode des plus proches voisins

2.1 Un modèle statistique

A l'heure actuelle, il est impossible de réaliser un modèle physique capable de prédire les avalanches sur des zones étendues. En effet, les propriétés mécaniques du mélange neige - glace qui forme le manteau neigeux ne sont pas encore suffisamment connues. De plus, il est impossible de modéliser la formation de congères. Finalement, il faudrait pouvoir disposer des mesures pour chaque site, ce qui est financièrement et humainement impossible.

Comparés aux modèles physiques, les modèles statistiques sont plus simples. Ils partent du principe que des conditions semblables entraînent des conséquences semblables. Les prévisions se reposent donc sur une base de données dans laquelle les paramètres significatifs sont enregistrés sur une période plus ou moins longue. Cette base de données est l'équivalent de la mémoire associative du prévisionniste.

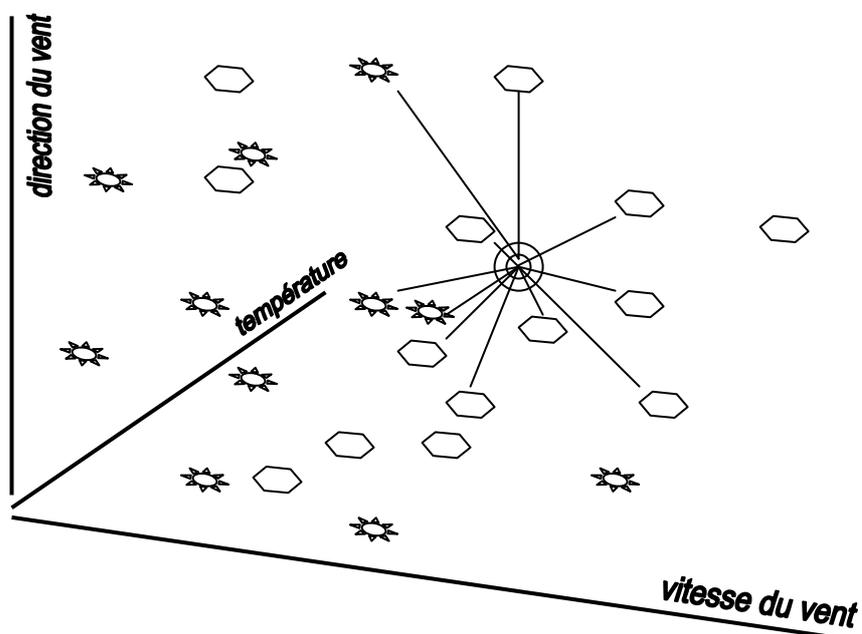
2.2 Principe général

La méthode des plus proches voisins a vu le jour en Suisse. Cet outil incontournable dans le domaine, permet aux prévisionnistes d'étayer leurs prévisions en comparant leurs informations à celles de journées antérieures similaires.

Les p variables enregistrées dans la base de données, définissent pour chaque jour i un vecteur X_i ($x_{i,k}$, $k=1$ à p) dans un espace d'état à p dimensions. Toute la difficulté réside alors dans la définition d'un critère de comparaison. Dans la méthode des plus proches voisins, celui-ci sera établi à partir de "distances" (euclidiennes) d_{ij} calculées entre les points représentatifs des journées i et j :

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik}^2 - x_{jk}^2)}$$

Pour chaque jour, les K plus proches voisins peuvent ainsi être identifiés. K sera choisi de manière judicieuse en fonction de la taille de la base de données. L'expérience montre que K=10 est un choix idéal car facilement affichable sur écran et aisément convertible en pourcentage.



*Figure 2
Représentation schématique en 3D de la méthode
des voisins les plus proches.*

*Les hexagones correspondent à des journées antérieures sans avalanches,
les étoiles à des journées avec avalanches.
Les dix voisins les plus proches sont identifiés.*

En réalité, on travaille dans un espace à 11 dimensions et l'on compte plus de 1500 jours enregistrés dans la base de données.

2.3 Mise à l'échelle (Scaling)

Jusqu'ici, les variables ont des unités ou des plages d'utilisation différentes. Par exemple, la variable "pénétration d'un pied dans la neige" varie entre 0 et 100 cm et la variable "température de l'air" se situe généralement dans la plage de -10°C à $+10^{\circ}\text{C}$. Si ces 2 variables étaient utilisées sans mise à l'échelle, une différence d'1cm pour l'une équivaldrait, dans l'espace euclidien, à une variation de 1°C pour l'autre. Ceci n'ayant pas de sens, il est nécessaire de convertir les données pour qu'elles s'ajustent dans des plages identiques allant par exemple de 0 à 100 (sans unités).

2.4 Pondération (Weighting)

Toutes les variables n'ont pas la même importance dans la physique des phénomènes. Aussi est-on amené à leur attribuer un poids relatif. Dans les premiers modèles, le poids était fixé par un spécialiste du massif montagneux auquel le modèle était destiné. Le scientifique procédait alors par essais et erreurs. Mais il s'est avéré que les spécialistes avaient une tendance toute naturelle à "overtuner" c'est-à-dire à privilégier dans leurs comparaisons les hivers récents. Ceci est une erreur puisque, a priori, l'hiver présent n'est pas plus ressemblant aux hivers récents qu'aux hivers lointains. Un modèle "overtuné" réagit donc très bien aux conditions étudiées ou prises en compte par les spécialistes, mais lorsque des cas plus éloignés se présentent, les performances du modèle se dégradent fortement.

2.5 Fitness, une mesure de la performance des modèles

Afin de pouvoir comparer les performances des différents modèles, il a fallu définir un critère de performance appelé "fitness". Ce critère a ses limites car il est simpliste de vouloir résumer la qualité de modèles, parfois complexes, en un seul nombre.

On effectuera donc la prévision "a posteriori" de chaque jour répertorié dans la base de données. Pour ce faire, on commence par rechercher pour chaque jour les 10 jours dont les caractéristiques enregistrées sont les plus proches. Ensuite, on établit une "prévision" de la façon suivante : si au moins 3 des 10 jours ont connu une avalanche, alors prévision = avalanche. En comparant les prédictions avec la réalité, on compte les jours d'avalanches correctement prédits (AvVrai), les jours d'avalanches non prédits (AvFaux), les

jours sans avalanche correctement prédits (NonAvVrai) et pour finir les fausses alertes (NonAvFaux). A partir de ces performances, on définit alors la Fitness :

$$\text{Fitness} = \frac{\text{AvVrai}}{\text{AvVrai} + \text{AvFaux}} + \frac{\text{NonAvVrai}}{\text{NonAvVrai} + \text{NonAvFaux}}$$

2.6 Sélection des paramètres.

Les paramètres sélectionnés dans le modèle doivent :

- représenter les conditions d'avalanches locales,
- être facilement mesurables tous les jours à la même place,
- être prédictibles au moins 24h à l'avance.

Cependant, parfois seules des raisons historiques guident le choix de paramètres enregistrés depuis des années.

13 paramètres enregistrés dans la base de données utilisée ici, par exemple : la température de l'air, la vitesse du vent, la direction du vent, la profondeur de pénétration des pas, la température de la neige, etc. 11 des paramètres sont utilisés pour la comparaison des jours dans l'espace euclidien, les 2 derniers paramètres précisent le type de journée rencontrée (avalanche ou non et échelle des risques). Pour la zone de Lochaber (région étudiée et siège du SAIS) les données sont collectées depuis 1991.

3 Algorithme génétique

3.1 Utilisation.

Comme énoncé plus haut, la pondération constitue une étape essentielle de la mise en place d'un modèle. En Ecosse, le dernier modèle programmé pour la prédiction des avalanches utilise une méthode inédite jusque-là : un algorithme génétique. En partant du constat qu'un spécialiste aura tendance à "overtuner" le système et que, par ailleurs, il coûte cher, il a été décidé de faire appel à cette forme d'intelligence artificielle basée sur le principe de

l'évolution de Darwin : “au fil des générations, les individus évoluent et seuls les meilleurs restent”.

L'algorithme génétique permet de trouver rapidement une solution satisfaisante même si celle-ci n'est pas toujours la meilleure. Ceci n'est toutefois pas grave car la meilleure solution est très souvent une solution “overtunée”.

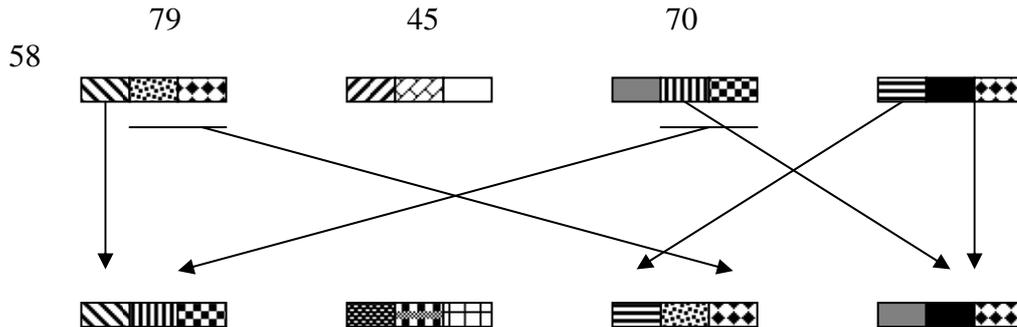
3.2 Un algorithme génétique utilisé avec la méthode des plus proches voisins.

L'algorithme est utilisé ici pour déterminer la combinaison des “poids” des paramètres qui donne les meilleures performances (mesurées ici par un indicateur appelé “fitness”). Dans la figure 3, chaque motif rectangulaire, appelé gène, représente le poids d'un paramètre, et l'association de trois poids (onze dans l'algorithme) forme une combinaison possible, appelée dans le jargon génétique, un individu ou aussi un chromosome. Chaque chromosome est alors caractérisé par sa fitness. Ce score traduit les performances du modèle pour le chromosome considéré. Après avoir ainsi calculé la qualité de chaque chromosome, le logiciel crée une nouvelle génération en modifiant les pondérations, dans l'espoir d'obtenir des individus de meilleure qualité après plusieurs générations. Les individus de la nouvelle génération seront le résultat de différentes sortes de “mutations” à partir de la génération précédente et de l'introduction aléatoire de nouveaux individus. Cette évolution aléatoire permettra, au fil des générations, de converger vers une solution optimale. Pour être sûr de ne pas écarter de bonnes solutions, on introduit un élément “élitiste” c'est-à-dire, on conserve systématiquement le meilleur individu de chaque génération. Il existe cependant une faible probabilité que des gènes évoluent dans une direction aléatoire.

Cette démarche, codée sous forme d'algorithme, est analogue au processus biologique et permet d'optimiser rapidement de nombreux problèmes. Il faut toutefois garder à l'esprit que tant que toutes les possibilités n'ont pas été essayées, on ne peut être sûr d'avoir identifié la meilleure.

Cette façon de faire n'est donc en rien comparable à un ajustement manuel qui risque d'“overtuner” le système pour les jours déjà présents dans la base de données et d'amoinrir les performances pour la prédiction de conditions inconnues. Malgré l'absence totale d'expertise dans le domaine climatique et nivologique, l'algorithme génétique a démontré son efficacité dans la prédiction d'avalanches par la méthode des plus proches voisins.

*Population après un certain nombre d'itérations ;
pour chaque individu une "fitness" peut être calculée.*



*Les meilleurs individus sont combinés pour l'itération suivante.
Un nouvel individu est introduit à la place du moins bon.*

Figure 3

Ce diagramme montre une seule itération d'un algorithme génétique appliqué à une population restreinte (4 individus) où les chromosomes sont formés de 3 gènes seulement. L'élitisme n'y est pas représenté.

4. Améliorations

4.1 Utilisation d'échelles non linéaires

De nombreux scientifiques ayant travaillé sur des modèles des plus proches voisins pour la prédiction des avalanches suggèrent d'introduire des échelles non linéaires pour chaque paramètre. En effet, on peut par exemple comprendre qu'une différence de température de 1°C soit plus significative autour de 0°C qu'autour de -10°C. Malheureusement, l'analyse statistique des données n'a pas été concluante dans cette voie. Une autre direction, plus prometteuse a donc été privilégiée.

4.2 Utilisation des données sur plusieurs jours

Méthode

Certains modèles ont démontré que la production d'une avalanche dépend des valeurs présentes mais également passées (jusqu'à 3 jours) de certains paramètres. Afin de tenir compte de cette "mémoire", l'espace d'état a été étendu à 44 dimensions (11 fois 4 jours). Comme pour le modèle original, un poids doit être attribué à chacune des 44 dimensions. Or il est humainement impossible d'explorer l'ensemble des solutions ; c'est donc à un algorithme génétique que le soin a été laissé de rechercher la pondération optimale. Il est intéressant de constater que les poids ainsi trouvés sont proches des valeurs utilisées dans un modèle italien et trouvées à l'aide d'autres considérations.

Vérification

Ce nouveau modèle permet de prédire correctement 84,5% des jours contre 81% pour l'ancien. Mais il a fallu vérifier que l'introduction des nombreux nouveaux paramètres (33), n'avait pas "overtuné" le modèle. Pour cela, on a divisé la base de données en deux moitiés de manière aléatoire ou temporelle. Ensuite, on a optimisé le modèle sur la première moitié (ensemble d'apprentissage) puis on l'a vérifié sur la seconde moitié (ensemble test dont les éléments jouent le rôle de situations futures inconnues du logiciel).

Les diagrammes de la figure 4 montrent l'évolution des modèles en cours d'optimisation. Chaque petite croix représente les performances (fitness) du modèle à un stade donné de l'apprentissage, d'abord sur l'ensemble d'apprentissage, puis sur l'ensemble de test. Ces résultats montrent que le nouveau modèle est moins enclin à "overtuning" que l'original. En effet, pour l'ancien modèle, passé un certain seuil de performance sur l'ensemble d'apprentissage, les performances sur l'ensemble de test n'augmentent plus. Le nouveau modèle ne se comporte pas ainsi car il extrait mieux la réalité physique cachée derrière les données statistiques.

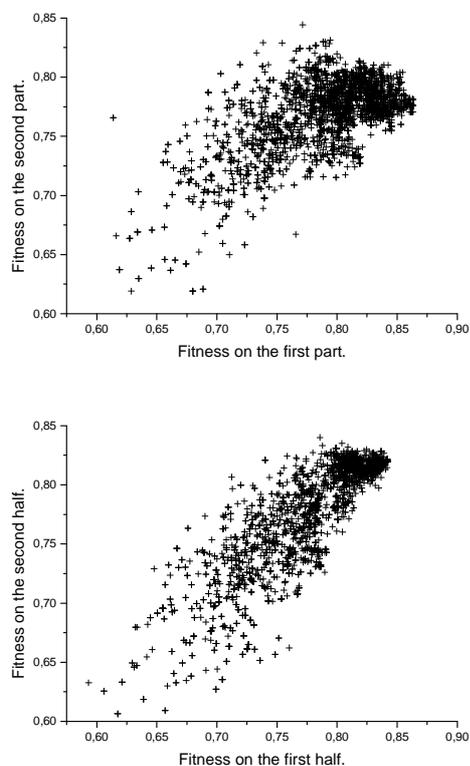


Figure 4

En haut, la réponse du premier modèle, en bas, celle du deuxième modèle.

Conclusions

La méthode des plus proches voisins, exploitée à l'aide d'un algorithme génétique, semble très bien convenir pour la prédiction d'avalanches.

Malgré les évidences, l'introduction d'échelles non linéaires ne semble pas apporter d'amélioration significative au modèle, mais des recherches sont encore à réaliser dans cette direction.

Par contre, l'utilisation des données sur les jours précédant le jour de prédiction permet d'augmenter significativement la qualité du modèle car il est dès lors possible de prédire correctement la nature de 84,5% des jours. Cela

correspond à une amélioration de 3% et constitue une amélioration significative à ce niveau de performance. En outre, le nouveau modèle proposé est moins enclin à “overtuning”.

Références bibliographiques

BARTON, B., Wright, B., *A chance in a million*, 2nd Edition, Scottish Mountaineering Trust, ISBN : 0-90-752159-2, 2000.

BOIS P., OBLED C. and GOOD W., (1975), *Multivariate data analysis as a tool for day-by-day avalanche forecast*, Snow Mechanics-Symposium, IAHS, Publication no 114.

BUSER O., BÜTLER, M., GOOD, W., *Avalanche forecast by the nearest neighbour method*, Avalanche Formation, Movement and Effect, IAHS, Publication, no 162, 1987.

BUSER O., *Two years experience of operational avalanche forecasting using the nearest neighbours method*, Annals of Glaciology, 13[31-34], International Glaciological Society, 1989.

FÖHN P., GOOD W., BOIS P. and OBLED C., *Evaluation and comparison of statistical and conventional methods of forecasting avalanche hazard*, Journal of Glaciology, Vol. 19, No. 81, 1977, p375 à 387.

FLANAGAN D., *Java in a Nutshell*, O'Reilly, 4th Edition, Springer-Verlag Berlin Heidelberg New York, ISBN : 0-596-00283-1, March 2002.

GASSNER M., ETTER H.-J., BIRKELAND Karl, *NXD2000 : An improved avalanche forecasting program based on the nearest neighbour method*.

HEIERLI J., PURVES R. S., FELBER A. and KOWALSKI J., *Verification of nearest neighbours interpretations in avalanche forecasting*, not yet published.

MICHALEWICZ, Fogel, *How to solve It : Modern Heuristics*, Springer-Verlag Berlin Heidelberg New York , ISBN : 3-540-66061-5, 2000.

PURVES R.S., SANDERSON M., *A methodology to allow avalanche forecasting on an information retrieval system*, University of Glasgow[Online]. http://dis.shef.ac.uk/mark/cv/publications/papers/my_papers/JDoc98.pdf (accessed in Mai 2004).

PURVES R.S., MORRISON K.W., MOSS G. and WRIGHT D.S.B.,(2003), *Nearest neighbours for avalanche forecasting in Scotland-development, verification and optimisation of a model*, Cold Regions Science and Technology 37 [344-355].

PURVES R.S., MORRISON K.W. and WRIGHT D.S.B., *Avalanche Forecasting Experiments for Torridon : A Feasibility Study for SportScotland*, [SAIS internal report], p 40.

PURVES R.S., *Data and it's use in Avalanche Forecasting Models*,[SAIS internal report].

PURVES R.S., Private Communication, 2004.