

Cyber-météorologiste

Alarmes de coups de vent à l'aide de programmation génétique

Le travail présenté vise à procurer au météorologiste un soutien automatisé à la prévision à court terme des coups de vent sur les lacs et les aérodromes suisses. Il fait appel à une méthode d'intelligence artificielle inspirée de la biologie: la programmation génétique.

Introduction

Dans la pratique qui a prévalu jusqu'à ce jour, 60 "objets d'alarme", lacs ou aérodromes, doivent être suivis en continu par les météorologistes de MétéoSuisse, et alarmés lorsqu'une situation de coups de vent ou de tempête se présente. Fondant leur décision sur l'ensemble des données qui leur sont disponibles, les météorologistes déclenchent "à l'oeil et à la main" des alarmes ponctuelles en cas d'occurrence probable d'un coup de vent de plus de 25 nœuds en l'un des 60 lieux. Soumis à une pression constante dans leur activité journalière, gérant des systèmes informatiques complexes, ils sont également affectés à des tâches de communication, rédigeant des textes ou des bulletins codés, répondant à d'innombrables téléphones. Du point de vue de l'ergonomie, il est par ailleurs connu qu'un humain, dès lors qu'il est scotché derrière un écran dans une activité de surveillance complexe, s'avère peu efficace lorsque la procédure requiert une attention implacable dans la durée.

C'est de la convergence de ces deux sources de tension que résulte l'idée d'un soutien automatisé au météorologiste. Fonctionnant en arrière-plan, le système n'attire son attention qu'en cas d'occurrence probable d'une forte rafale en l'un des 60 lieux sous surveillance. Averti, le météorologiste peut dès lors brièvement considérer le cas, décider de déclencher l'alarme, ou au contraire refuser la proposition du système.

L'idée

La programmation génétique repose sur le concept de la sélection naturelle, tel qu'il est connu en biologie. Il consiste à faire évoluer une société d'individus sur un grand nombre de générations en soumettant ces individus à une contrainte d'optimalité, tout en conservant au fil des générations les individus satisfaisant au mieux cette contrainte. Un dispositif de mutation et de croisement entre individus assure d'une génération à la suivante d'une part la variabilité entre les caractères des individus - garante de solutions inédites - d'autre part la conservation des caractères utiles. Les individus émergeant au terme du processus d'évolution représentent des solutions si possible optimales au problème donné.

L'application de cette méthode à la prévision météorologique à court terme a été étudiée dans le cadre de travaux de Bachelor conduits entre le Laboratoire d'intelligence artificielle de l'Université de Zurich [références] et MétéoSuisse. Les travaux ont ensuite été poursuivis en interne à Météo Suisse sous l'égide des auteurs. L'ensemble est conduit dans le cadre du projet OWARNA (Optimierung des Warnungsabläufe, Optimisation des processus d'alarme) actuellement en cours à MétéoSuisse.

L'apprentissage génétique

Dans notre cas, les "individus" évoqués plus haut sont de petits programmes informatiques (méthodes Java¹), aptes à produire des prévisions de vent à court terme en un lieu prédéterminé (lac ou aéroport). Pour ce faire, les programmes combinent des données pertinentes d'observation météorologiques et de prévisions numériques livrées par le modèle COSMO-2. Le modèle COSMO-2 est décrit dans "Le bol d'or virtuel", article paru dans le numéro 12 de Météo-magazine.

Nos petits programmes n'ont bien entendu d'autre existence que virtuelle dans la mémoire d'un ordinateur. Ils peuvent être décrits par leur phénotype, à savoir leur corps, dans notre cas le code de chacun des programmes, ou par leur génotype, une description abstraite du code sur laquelle la sélection opère. A titre de comparaison, le génotype d'un être vivant est le message le décrivant, codé dans son ADN, le phénotype est l'être lui-même.

A chaque génération, la capacité des programmes à prédire les rafales de vent est vérifiée, et les meilleurs d'entre eux sont sélectionnés. Des processus de mutation et de croisement sont alors engagés. Opérant uniquement sur les gagnants, ils livrent une nouvelle génération de programmes. Ces étapes de vérification, sélection et mutation/croisement sont répétées au fil des générations jusqu'à un nombre maximal, prédéfini. Dans notre implémentation, une génération est composée de 1000 individus évoluant au cours de 500 générations. A chaque génération, seuls les 400 meilleurs individus se reproduisent et délivrent les 1000 individus de la génération suivante.

Cette phase d'évolution des programmes informatiques, appelée aussi phase d'apprentissage, n'a lieu qu'une seule fois pour chaque lac ou aéroport et se base sur des données historiques. Au terme de l'apprentissage, après 500 générations, seuls sont conservés les 20 meilleurs individus, qui sont, on le rappelle de petits programmes, méthodes java, livrant chacune une prévision de rafale à court terme pour cet objet.

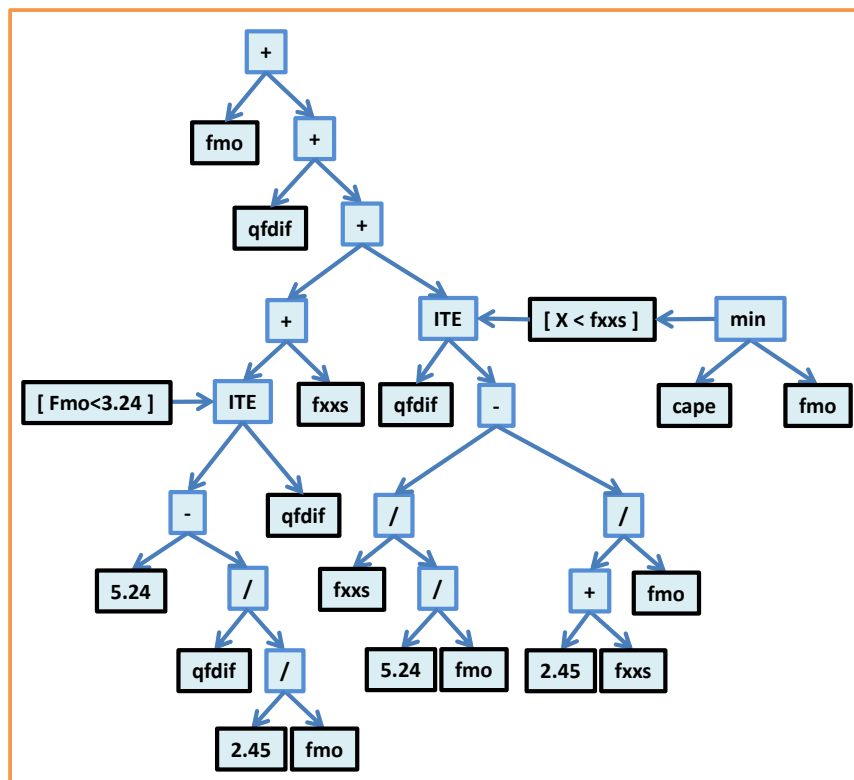


Figure 1. Exemple du génotype d'un algorithme développé par le processus de sélection naturelle pour l'aéroport de Granges, sur le plateau suisse.

Sa topologie, développée au cours du processus d'évolution, est celle d'un arbre. Les nœuds sont des opérateurs, les feuilles décrivent les éléments d'information émanant soit des systèmes d'observation, soit du modèle numérique de prévision COSMO-2.

¹ Java est un langage de programmation

Une notion de “dynamique des essais” émerge ici. De manière générale, la sélection naturelle ne livre pas un, mais un ensemble d’individus, tous quelque peu différents, satisfaisant cependant tous, dans un sens propre à chacun d’entre eux, la contrainte optimale fixée. Ainsi, chaque individu livre une réponse différenciée à une question posée à tous, et l’essaim produit une réponse collective pouvant être interprétée dans un sens probabiliste. Dans notre cas, les 20 méthodes Java associées à chaque objet (lac ou aéroport) génèrent ensemble une probabilité d’occurrence d’une rafale supérieure à 25 noeuds sur cet objet, dans les trois heures suivantes. Un seuil optimal de probabilité d’occurrence, noté Q^* sur la figure 2, est également livré à la suite d’une phase de vérification. Pouvant satisfaire divers critères, il peut par exemple être déterminé en prenant en compte le profil de risque d’un utilisateur, en s’inspirant de la méthode décrite dans “Le dilemme de Cassandra,” paru dans le numéro 11 de Météo-magazine. La partie supérieure de la figure 2 illustre la suite des opérations décrites jusqu’ici.

En routine opérationnelle

En mode opérationnel, la probabilité d’occurrence d’une rafale, notée P sur la figure 2, est calculée chaque 10 minutes sur la base des dernières données d’observation et de prévision disponibles. Une proposition d’alarme est activée si cette probabilité excède le seuil optimal d’alarme Q^* déterminé lors de l’apprentissage. La partie inférieure de la figure 2 décrit ces différentes étapes.

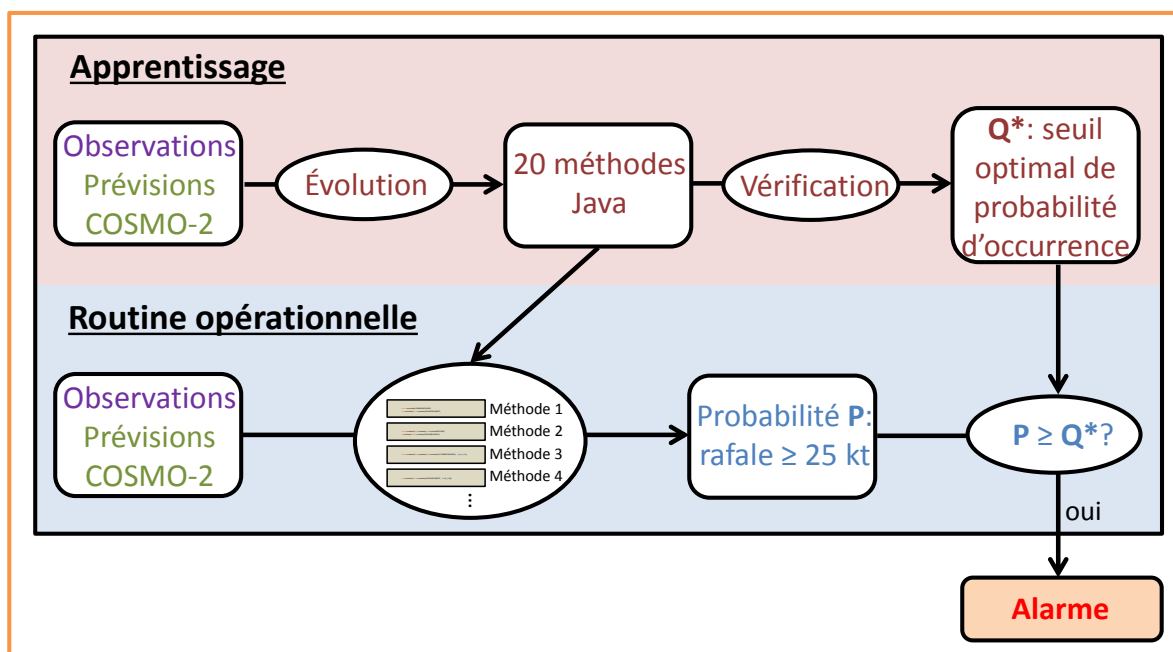


Figure 2. Le panneau supérieur décrit la phase d’apprentissage, exécutée une fois pour chaque objet d’alarme sur la base de données historiques. Le résultat de cet apprentissage est d’une part la liste des algorithmes développés, nommés ici méthodes Java (l’une d’entre elles est exhibée à la figure 1). L’apprentissage livre également un seuil optimal de probabilité au delà duquel une alarme doit être donnée. Aussi bien les méthodes que le seuil sont spécifiques à un objet, lac ou aéroport.

En mode opérationnel, les dernières informations disponibles, émanant aussi bien des systèmes d’observation que des prévisions numériques, sont présentées aux méthodes. Celles-ci livrent une probabilité d’occurrence de rafales de plus de 25 noeuds, qui est comparée au seuil optimal calculé lors de l’apprentissage. L’alarme est déclenchée si cette probabilité excède le seuil.

Résultats

Le système délivrant des alarmes, sa capacité prévisionnelle est exprimée dans les termes de son taux de fausses alarmes, et de son taux de détection. L'évaluation de ces taux fait appel à la méthode présentée dans l'article intitulé "Le Dilemme de Cassandre", publié dans le numéro 11 de Météo-magazine. La figure 3 décrit les résultats obtenus suivant cette méthode. Le taux de fausses alarmes, (FAR, False Alarm Ratio) est reporté en abscisse sur cette figure, le taux de détection (HR, Hit Rate) en ordonnée. La courbe figurée par les points rouges, appelée courbe ROC (Relative Operation Characteristic), est équipée d'une abscisse curviligne exprimée en termes de probabilités. L'aérodrome considéré est Granges (LSZG), sur le plateau Suisse. Un système d'alarme idéal se situerait dans le coin supérieur gauche du diagramme. Il détecterait tous les événements (HR=1) et ne délivrerait aucune fausse alarme (FAR=0).

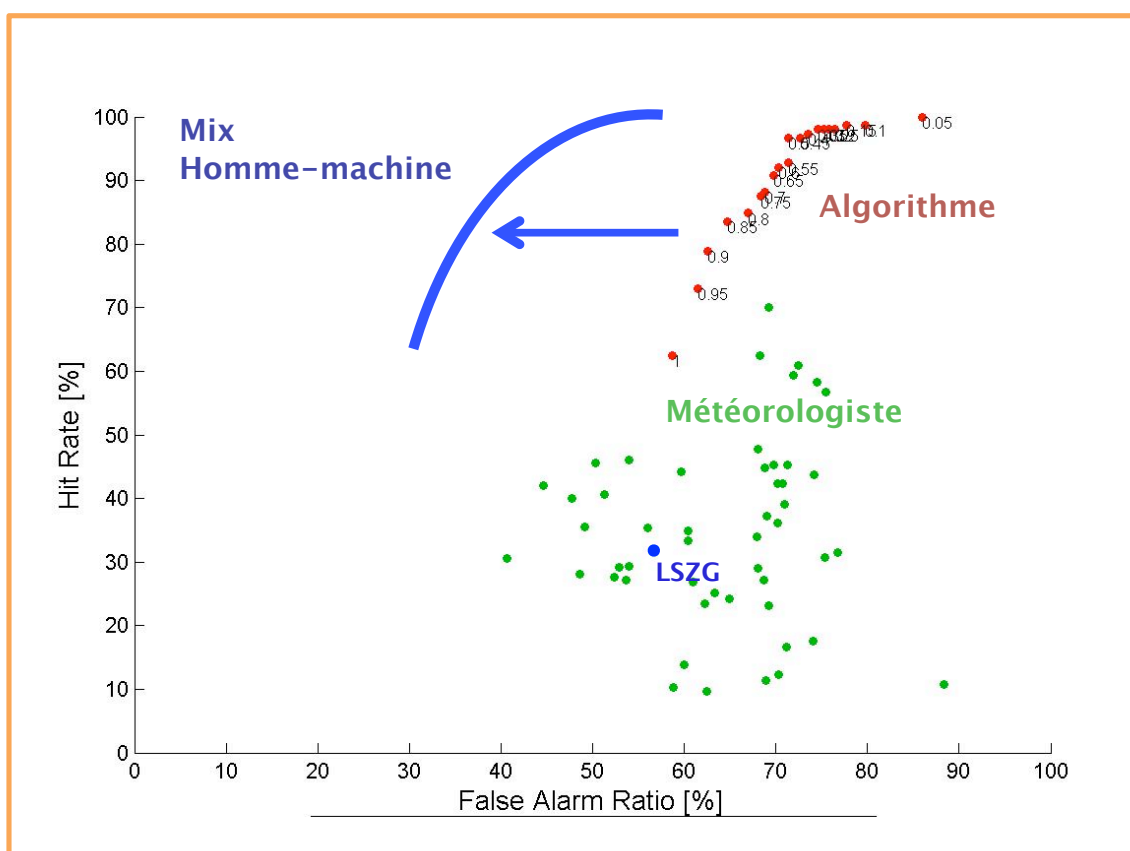


Figure 3. Diagramme exhibant les performances de divers systèmes d'alarmes pour des coups de vent égaux ou supérieurs à 25 noeuds (46 km/h). Pour un objet donné (lac ou aérodrome), la performance en cet objet est livrée par ses coordonnées: en abscisse le taux de fausses alarmes, en ordonnée le taux de détection. Le nuage des points verts représente les résultats obtenus par les météorologistes sur l'ensemble des objets. La séquence des points rouges représente le résultat (courbe ROC) de l'ensemble des algorithmes génétiques pour l'aérodrome de Granges (LSZG). Le taux de détection y est excellent, au prix d'un taux excessif de fausses alarmes. La symbiose des deux systèmes, homme et machine, devrait conduire au résultat esquissé en bleu.

Sur la figure 3, les points rouges exhibent la performance du système d'alarme automatique. Sous chaque point est inscrit le seuil de probabilité correspondant au résultat obtenu. Le point situé à l'extrémité inférieure gauche correspond au HR et au FAR obtenus si le système livre une alarme pour une probabilité d'occurrence de 100%. Aussi bien le taux de

fausses alarmes que celui de détection sont bas. Puis, par intervalles de 5%, on atteint finalement le point situé à l'extrémité supérieure droite de la ROC, correspondant au HR et FAR obtenus lorsque le système délivre ses alarmes à partir d'une probabilité d'occurrence (extrêmement basse) de 5% déjà. Le taux de détection est alors élevé, au prix d'un taux prohibitif de fausses alarmes. On observe que pour un seuil de probabilité de 50%, auquel les alarmes sont délivrées si la probabilité de leur occurrence égale ou dépasse ce seuil, le système détecte environ 95% des cas mais délivre des fausses alarmes à un taux de 70% environ.

L'essaim des points verts représente la qualité des alarmes actuelles lancées manuellement par les météorologistes. La probabilité de détection est généralement basse, souvent inférieure à 50%. Suivant l'introduction, la toute première rafale échappe souvent à l'attention du météorologiste occupé à toute une foule d'autres tâches. L'évènement est alors comptabilisé comme manqué, même si une alarme est lancée peu après. Ainsi, le système de propositions automatiques d'alarmes a l'avantage d'augmenter la probabilité de détection de la toute première rafale de vent, au prix d'un taux de fausses alarmes élevé. Dans la pratique cependant, on attend du météorologiste, grâce à sa longue expérience, qu'il filtre les fausses alarmes proposées par le système et diminue ainsi le taux de fausses alarmes délivrées, contribuant ainsi à la qualité générale du système combiné homme-machine.

Conclusion

Le système de propositions automatiques d'alarmes présenté apporte un soutien au météorologiste engagé dans son rôle de surveillance continue des conditions de vent en 60 lieux différents. Le système suit en continu, et avec une attention indéfectible, l'évolution météorologique aux 60 objets d'alarme et avise le météorologiste lorsque la situation devient potentiellement critique.

La mise en place de ce dispositif automatique conduira à une amélioration générale de la détection des toutes premières rafales, sur tout le territoire. En filtrant les alarmes proposées, le météorologiste contribuera à diminuer le taux de fausses alarmes. C'est grâce à la combinaison homme-machine qu'une amélioration substantielle de la qualité des alarmes de coups de vent sur les lacs et aérodromes régionaux devrait être réalisée en Suisse. Cette remarque optimiste nous a suggéré le titre de la contribution.

Références

- Universität Zürich, Laboratorium für künstliche Intelligenz:
<http://www.ifi.uzh.ch/ailab.html>
- Roman Voisard, 2010. Genetic programming, evolutionary strategies and model output statistics to predict visibility at Zurich airport. University of Zürich; Department of Informatics.
- Manuela Züger, 2010. Potential use of genetic programming in storm warning systems. Bachelor Thesis. University of Zürich; Department of Informatics.