Composition optimale d'équipes d'athlétisme

Frédéric Gardi

Bouygues e-lab, 32 avenue Hoche, 75008 Paris, France.

fgardi@bouygues.com

Dans cette note, nous présentons un problème d'optimisation combinatoire que l'on peut qualifier de "problème du quotidien". Celui-ci, bien qu'aux enjeux modestes, fait un excellent exercice de recherche opérationnelle pour les étudiants, ainsi qu'un sujet de vulgarisation à destination du grand public.

En 2003, Martin et Rottembourg [2] présentaient au colloque ROADEF l'approche qu'ils avaient mise en œuvre pour résoudre un problème de constitution d'équipes de natation. Ils concluaient leur présentation en mentionnant qu'un problème similaire se posait pour la constitution d'équipes d'athlétisme. En effet, chaque année les clubs d'athlétisme français se réunissent à différents niveaux pour une compétition appelée "interclubs". Le principe de la compétition est le suivant. Chaque club doit constituer une équipe masculine et une équipe féminine. Les équipes masculines s'affrontent sur 21 épreuves réparties en quatre groupes : les courses (100 m, 200 m, 400 m, 110 m haies, 400 m haies, 800 m, 1500 m, 3000 m, 5000 m, 3000 m steeple, 5000 m marche), les sauts (longueur, hauteur, triple saut, perche), les lancers (poids, disque, javelot, marteau), les relais (4 × 100 m, 4 × 400 m). De la même manière, les équipes féminines s'affrontent sur 19 épreuves semblables aux épreuves masculines (seuls le 5000 m et le 3000 m steeple n'ont pas d'équivalent chez les femmes).

Chaque club peut aligner au plus deux athlètes par épreuve (bien entendu, hommes et femmes concourent séparément). Chaque athlète peut concourir sur deux épreuves au plus appartenant à des groupes différents parmi courses, sauts et lancers. Chaque athlète peut participer en sus à une des deux épreuves de relais, ceux-ci se déroulant en toute fin de compétition. Les performances des athlètes sur chaque épreuve sont convertis en points à l'aide de tables (celles-ci, appelées "tables hongroises", sont établies par application d'une formule du type $p(x) = \lfloor ax^2 + bx + c \rfloor$ pour chaque épreuve, où $p(x)$ est le nombre de points obtenus pour une performance x exprimée dans l'unité adéquate). Ainsi, l'objectif pour un club est de maximiser son nombre total de points à l'issu de la compétition.

Les interclubs se jouent en deux rencontres. À l'issu de ces deux tours, les clubs sont classés à l'échelon national en fonction de leur performance (classement qui est regardé pour l'octroi de subventions). Cela en fait un rendez-vous crucial dans la vie des clubs d'athlétisme français, petits ou grands. De par notre expérience, nous pouvons affirmer que la composition des équipes d'interclubs est un casse-tête pour les entraîneurs d'athlétisme. Au-delà de la complexité même du problème d'optimisation qu'ils ont à résoudre (une cinquantaine d'athlètes à affecter à $21 + 19$ épreuves), ils doivent faire face à des impondérables, jusqu'au jour même de la compétition (absences, blessures). Un outil informatique d'aide à la décision leur permettant de composer, et surtout de recomposer, leur équipe de façon optimale en fonction de la disponibilité des athlètes, de leurs performances du moment, ou encore du planning de la compétition (par exemple, il est préférable de ne pas aligner un athlète sur 400 m, s'il a le saut à la perche à réaliser en même temps) serait donc le bienvenu.

Pour le chercheur opérationnel, la résolution du problème en question ne pose vraiment pas de difficulté, une fois observé que celui-ci peut être formulé comme un problème de flot de coût maximum lorsque la question des relais, qui n'est pas décisive, est laissée de côté. En effet, soit A l'ensemble des athlètes, G les différents groupes d'épreuves et E l'ensemble des épreuves hors relais. Le réseau est construit de la façon suivante. À chaque athlète, ainsi qu'à chaque épreuve, est associé un nœud. Le nœud origine est relié à chaque nœud "athlète" par un arc de capacité 2 et chaque nœud "épreuve" est reliée au nœud destination par un arc de capacité 2. Ensuite, à chaque athlète est associé $|G|$ nœuds représentant les groupes d'épreuves. Tout nœud "athlète" est relié à chacun de ses noeuds "groupe" par un arc de capacité 1 et tout nœud "groupe" associé à un athlète est relié à chaque nœud "épreuve" qui lui appartient par un arc de capacité 1 et de coût $P_{a,e}$ égal au nombre de points espéré si cette épreuve est affecté à l'athlète. Le réseau obtenu comporte $|A|(|G|+1)+|E|$ nœuds et $(|A|+|E|)(|G|+1)$ arcs. Essentiellement deux types d'algorithme ont été proposés pour calculer un flot maximum de coût minimum dans un réseau [4] : la méthode d'augmentation du flot (augmenter le flot le long de chemins de coût minimum) et la méthode de réduction du coût (calculer un flot maximum puis éliminer les circuits de coûts négatifs). Ces méthodes seront ici toutes deux efficaces tant en théorie (puisque la valeur du flot maximum est bornée par $2\max\{|A|,|E|\}$, les coûts sont positifs et le réseau est acyclique) qu'en pratique (puisque $|A| \leq 100$, $|G| = 3$, $|E| = 40$).

Toutefois, afin d'anticiper d'éventuelles évolutions rendant une modélisation par les flots impossible (par exemple, si l'on considère les contraintes d'exclusion mutuelle entre épreuves dues au planning de la compétition), nous avons opté pour une approche par programmation linéaire en nombres entiers. Le programme peut être écrit comme suit, où $x_{a,e}$ est égal à 1 si l'athlète $a \in A$ est affecté à l'épreuve $e \in E$ et à 0 sinon : maximiser $\sum_{a\in A, e\in E} P_{a,e}x_{a,e}$, sujet à $\sum_{a\in A} x_{a,e} \leq 2$ pour tout $e \in E$ (au plus deux athlètes par épreuve), $\sum_{e\in E} x_{a,e} \leq 2$ pour tout $a \in A$ (au plus deux épreuves par athlète), $\sum_{e\in g} x_{a,e} \leq 1$ pour tout $a \in A$ et $g \in G$ (au plus une épreuve de chaque groupe par athlète). On peut montrer que la matrice de ce programme est totalement unimodulaire, impliquant que toute base optimale de sa relaxation linéaire est entière [3]. En pratique, la résolution de ce programme linéaire par l'algorithme du simplexe implémenté dans la librairie GLPK [1] est instantanée (une centaine d'itérations de simplexe), d'autant plus qu'il est facile de lui fournir une bonne base réalisable (à partir d'une affectation heuristique des athlètes aux épreuves).

Notre logiciel, qui se présente sous la forme d'un classeur Excel accompagné d'une DLL dédiée à la résolution du problème d'optimisation, a été utilisé en 2008 pour composer les équipes d'interclubs homme et femme du BCI Athlétisme (Isle-sur-la-Sorgue, Vaucluse).

Références

1. A. Makhorin (2007). GLPK : Gnu Linear Programming Kit, version 4.24. `http://www.gnu.org/software/glpk/`
2. B. Martin, B. Rottembourg (2003). Constitution optimale d'équipes de natation. In *ROADEF 2003*. Avignon, France.
3. A. Schrijver (2003). *Combinatorial Optimization : Polyhedra and Efficiency.* Algorithms and Combinatorics 24, Vol. A, Springer-Verlag, Berlin, Allemagne.
4. R.E. Tarjan (1983). *Data Structures and Network Algorithms.* CBMS-NSF Regional Conference Series in Applied Mathematics 44, SIAM Publications, Philadelphie, PA.