Les mathématiques de l'OPR – introduction

Original: *The Blue Alliance Blog* *The Math Behind OPR — An Introduction* Eugene Fang, octobre 2017

Le *coefficient de puissance offensive* : OPR (*Offensive Power Rating*) est une statistique souvent utilisée dans la communauté *FIRST* pour comparer les performances des équipes sur le terrain. Il est aussi courant d'entendre des choses comme « *ce jeu est un mauvais jeu pour l'OPR* » ou « *l'OPR est un indicateur de performance assez précis cette année* ». Pour comprendre les forces et les limites de l'OPR, les maths sont essentielles. Cet article fournira une introduction aux mathématiques derrière l'OPR. Un futur article se penchera sur l'effet du design du jeu sur l'utilité de l'OPR en tant que statistique.

Note: Bien que cet article soit principalement destiné aux personnes qui ont au moins étudié Algèbre 1, l'expérience des matrices aidera beaucoup. Quoi qu'il en soit, j'espère que cet article mettra en lumière le fait que les matrices sont beaucoup plus qu'un moyen intelligent d'organiser des nombres. Soyez attentif dans vos futures classes d'algèbre linéaire!

Introduction

L'OPR est un terme plus largement utilisé que ce que Karthik Kanagasabapathy de l'équipe 1114 appelait autrefois la « contribution calculée ». L'OPR tente de calculer la contribution d'une équipe au score final de son alliance. S'il n'y avait pas d'alliances en Compétition de robotique *FIRST*, la contribution calculée d'une équipe serait simplement son score de match moyen. C'est une métrique facile et simple à calculer. Cependant, l'utilisation de scores moyens avec des alliances, tout en générant un indicateur correcte, n'est pas sans risque car un score moyen est une combinaison de performances des trois équipes sur l'alliance. L'équipe 1114 a commencé à utiliser l'OPR en 2004 parce qu'ils voulaient une mesure plus précise que le score moyen pour évaluer comment chaque équipe avait joué. Ils cherchaient un moyen de déterminer ce que chaque équipe contribuait en moyenne au score de son alliance. Ici, l'OPR fait une hypothèse clé: le score final d'une alliance est une combinaison linéaire de la contribution individuelle de chaque membre de l'alliance. En d'autres termes, l'addition des points gagnés par chaque membre d'une alliance de trois équipes donne le score final de l'alliance. Cela ressemble à:

$$A + B + C = score$$

Ici, A, B et C sont les contributions individuelles de chaque équipe au score d'alliance. Mais comment calculons-nous A, B et C? La réponse est la suivante: résoudre un système d'équations.

Un exemple simple

Pour simplifier les choses, imaginons que les alliances soient composées de 2 équipes au lieu de 3. En outre, concentrons-nous uniquement sur une alliance dans un match. Disons que les équipes A et B ont marqué 10 points ensemble. Cela peut être décrit par l'équation:

$$A + B = 10$$

Nous ne pouvons toujours pas calculer combien de points chaque équipe a marqué. A a marqué 8 et B a marqué 2? Est-ce plutôt que A a marqué 10 et B a marqué 0? Les deux équipes ont-elles marqué 5 points chacune? Actuellement, nous avons un système sous-déterminé avec plus de variables (équipes) que d'équations (résultats de match). Nous avons besoin de plus d'informations pour trouver des valeurs pour toutes les variables. Heureusement, au fur et à mesure de la compétition, nous avons plus de matches avec lesquels travailler. Disons que plus de matchs sont joués avec les résultats suivants:

$$A + C = 13$$
$$B + C = 7$$

Maintenant nous avons 3 équations et 3 inconnues - nous devrions pouvoir résoudre A, B et C! Avec une algèbre de base (ou un ordinateur pour les résoudre), nous trouvons que la contribution calculée (OPR) de l'équipe A est de 8 points, B de 2 et C de 5. Vous savez maintenant comment calculer l'OPR! Mais attendez, il y a quelques problèmes...

Un système surdéterminé

Essayons un autre exemple avec plus de matchs:

$$A + B = 10$$
$$A + C = 13$$
$$B + C = 7$$
$$A + D = 15$$

Avec 4 équations et 4 inconnues, nous pouvons résoudre cela comme tantôt. La solution donne A = 8, B = 2, C = 5 et D = 7. Cependant, cela ne continuera à fonctionner tandis que plus d'équations seront ajoutées que si toutes les équipes sont parfaitement cohérentes et si les équipes ont marqué le même nombre de points dans leurs matchs respectifs. Mais en Compétition de robotique *FIRST*, les robots se cassent, les équipes sont irrégulières, et parfois le jeu lui-même peut être le problème (mais c'est un sujet pour un autre article). Ajoutons le résultat de match suivant:

$$B + D = 10$$

Tenter de résoudre ce système d'équations ne donne aucune solution. Maintenant nous avons un système surdéterminé avec 5 équations et 4 inconnues - aucune valeur de A, B, C et D ne peut simultanément satisfaire toutes les équations (en général). Au lieu de trouver des valeurs qui satisfont parfaitement toutes les équations, nous ne devons maintenant trouver une solution approximative qui minimise l'erreur dans toutes les équations. Pour ce faire, nous nous tournons vers l'algèbre matricielle.

Équations en tant que matrices

Revenons à notre exemple simple. Il peut être réécrit sous forme de matrice :

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 10 \\ 13 \\ 7 \end{bmatrix}$$

En appliquant la multiplication matricielle, nous pouvons voir que la forme matricielle est la même que notre ensemble original d'équations:

$$1A + 1B + 0C = 10$$
$$1A + 0B + 1C = 13$$
$$0A + 1B + 1C = 7$$

Si nous avons '*n*' équipes et '*m*' matchs (souvenez-vous, nous ne regardons toujours qu'une seule alliance par match), le système d'équations prend la forme de $M x = s$. Ici, M est une matrice m par n où chaque rangée indique quelles équipes étaient dans une alliance lors d'un match donné (il ne devrait y avoir que des 1 et 0 : 1 si une équipe était dans le match et 0 si elle ne l'était pas). x est le vecteur m par 1 des OPR (variables que nous essayons de résoudre). Et s est un vecteur m par 1 des scores d'alliance.

Une façon de résoudre pour x est de multiplier-par-la-gauche les deux côtés par M^{-1}, l'inverse de la matrice M. En pratique, on ne le ferait pas nécessairement. L'inverse de M n'existe pas toujours, et il existe d'autres manières plus rapides et plus stables numériquement (robustes aux erreurs d'arrondissement d'ordinateur) pour résoudre x surtout quand M devient très grand (beaucoup d'équipes, beaucoup de matchs). Cependant, nous l'utilisons ici pour démontrer que nous obtenons la même solution qu'auparavant. En appliquant cette inversion, notre équation matricielle devient :

$$M^{-1}Mx = M^{-1}s$$

$$x = M^{-1}s$$

$$x = \begin{bmatrix} 0.5 & 0.5 & -0.5 \\ 0.5 & -0.5 & 0.5 \\ -0.5 & 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 10 \\ 13 \\ 7 \end{bmatrix}$$

En multipliant, on voit que la solution est la même que précédemment: A = 8, B = 2, C = 5.

$$x = \begin{bmatrix} 0.5 * 10 + 0.5 * 13 - 0.5 * 7 \\ 0.5 * 10 - 0.5 * 13 + 0.5 * 7 \\ -0.5 * 10 + 0.5 * 13 + 0.5 * 7 \end{bmatrix} = \begin{bmatrix} 8 \\ 2 \\ 5 \end{bmatrix}$$

Solution des moindres carrés

Génial, maintenant nous pouvons résoudre les équations sous forme de matrice. Mais comment cela nous aide-t-il lorsque le système est surdéterminé? Jetons un coup d'œil à notre système surdéterminé:

$$A + B = 10$$
$$A + C = 13$$
$$B + C = 7$$
$$A + D = 15$$
$$B + D = 10$$

Sous forme matricielle :

$$Mx = s$$

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} x = \begin{bmatrix} 10 \\ 13 \\ 7 \\ 15 \\ 10 \end{bmatrix}$$

Comme avant, nous ne pouvons pas résoudre ce système d'équations. Pour remédier à cela, nous pouvons multiplier les deux côtés par la transposition de M et réessayer.

$$M^T Mx = M^T s$$

Lorsque nous avons multiplié-par-la-gauche les deux côtés du système surdéterminé ($Mx = s$) par la transposition de la matrice binaire (M^T), nous créons ce qu'on appelle «l'équation normale». La solution de l'équation normale est la « solution des moindres carrés » du système surdéterminé. Il y a une preuve pour cela, mais c'est hors propos ici.

Voyons ce que ça donne:

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} x = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 10 \\ 13 \\ 7 \\ 15 \\ 10 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 1 & 1 & 1 \\ 1 & 3 & 1 & 1 \\ 1 & 1 & 2 & 0 \\ 1 & 1 & 0 & 2 \end{bmatrix} x = \begin{bmatrix} 38 \\ 27 \\ 20 \\ 25 \end{bmatrix}$$

Maintenant, nous pouvons résoudre pour x! Pour plus de simplicité, convertissons ceci en équations séparées et résolvons.

$$3A + 1B + 1C + 1D = 38$$
$$1A + 3B + 1C + 1D = 27$$
$$1A + 1B + 2C + 0D = 20$$
$$1A + 1B + 0C + 2D = 25$$

Voici les OPR! A=7,75; B=2,25; C=5; D=7,5

Ok, qu'est-ce que la solution des moindres carrés?

Nous venons de résoudre l'équation normale, ce qui signifie que le résultat est la solution des moindres carrés du problème original. Comme il n'y avait pas de solution exacte au système surdéterminé, nous avons plutôt résolu pour la meilleure approximation de la solution. La « meilleure » solution peut être définie de différentes manières (par exemple écart minimum absolu, erreur quadratique moyenne minimale, etc.). Pour l'OPR, les moindres carrés définissent de facto « meilleur ». En d'autres termes, nous avons trouvé la solution (valeurs pour A, B, C et D) qui minimise la somme des erreurs au carré dans toutes nos équations. Regardons à nouveau nos équations originales :

$$A + B = 10$$
$$A + C = 13$$
$$B + C = 7$$

$$A + D = 15$$
$$B + D = 10$$

Avec nos solutions, on obtient:

$$7.75 + 2.25 = 10 + error$$
$$7.75 + 5 = 13 + error$$
$$2.25 + 5 = 7 + error$$
$$7.75 + 7.5 = 15 + error$$
$$2.25 + 7.5 = 10 + error$$

Les équations ne sont plus *parfaites* - elles sont erronnées (0, -0,25, 0,25, 0,25 et -0,25 respectivement). Si nous additionnons le carré de ces erreurs, nous obtenons:

$$0^2 + (-0.25)^2 + (0.25)^2 + (0.25)^2 + (-0.25)^2 = 0.25$$

0,25 est le minimum pour cette somme d'erreurs au carré. Choisissez n'importe quelle autre valeur pour A, B, C et D, et la somme sera plus grande. Par exemple, choisissons A = 7, B = 3, C = 5, D = 7,5:

$$7 + 3 = 10 + error$$
$$7 + 5 = 13 + error$$
$$3 + 5 = 7 + error$$
$$7 + 7.5 = 15 + error$$
$$3 + 7.5 = 10 + error$$

Maintenant, nos erreurs sont: 0, -1, 1, -0,5 et 0,5 respectivement. En calculant la somme du carré de ces erreurs, on obtient:

$$0^2 + (-1)^2 + (1)^2 + (-0.5)^2 + (0.5)^2 = 2.5$$

Ces nouvelles valeurs pour A, B, C et D sont pires que notre solution optimale des moindres carrés; nous obtenons une somme plus élevée d'erreurs au carré: 2,5 contre 0,25.

OPR vs score moyen

En revenant à l'objectif original de trouver une mesure qui traduit plus fidèlement la contribution individuelle d'une équipe que son score final moyen, nous voyons que l'OPR tente de séparer mathématiquement la contribution d'une équipe de la production globale de l'alliance. Un fait intéressant à propos de l'OPR est que si vous prenez l'OPR moyen de toutes les équipes en tournoi régional, le résultat devrait être égal au score moyen de toutes les équipes de l'événement divisé par 3. Cela est logique car trois équipes avec un OPR moyen devraient se combiner pour donner un score moyen de match. Un autre point à considérer, il est possible d'avoir des OPR négatifs. Cela signifie qu'en moyenne, la présence d'une équipe sur le terrain fait baisser le score de l'alliance. Ceci est rare mais survient de temps en temps, surtout quand les pénalités font baisser votre propre score au lieu d'augmenter le score de votre adversaire.

Dernières remarques

Ici, nous avons simplifié le problème en considérant seulement une alliance par match, avec deux équipes sur l'alliance. Pour les matchs réels 3 contre 3, cela ne change que quelques choses (mais rien des maths). Au lieu que *M* et *s* ayant le même nombre de lignes que le nombre de matchs, ils auraient le double des lignes (une pour chaque alliance du match). De plus, la matrice *M* aurait toujours trois 1 par rangée au lieu de deux, puisque trois équipes jouent.

Pour de très grands systèmes d'équations (par exemple en calculant « OPR Championnat » en utilisant les quelques 15 000 matchs d'une saison), l'utilisation de sous-matrices peut accélérer considérablement le calcul.

Conclusion

À la base, l'OPR ne fait que résoudre un système d'équations – tel qu'enseigné dans une classe d'algèbre de base. Cependant, lorsqu'il y a plus d'équations que de variables (plus de matchs d'alliances que le nombre d'équipes), le système devient en général surdéterminé. L'OPR trouve la solution des moindres carrés à ce système surdéterminé en trouvant une solution à la forme d'équation normale du système.

Un grand merci à Karthik et Ether pour leur contribution pour cet article.

PS: Jaci Brunning a posté un excellent article qui approfondit l'algèbre linéaire qui a été présentée ici.