IA04 – Systèmes Multi-Agents

Théorie du vote : résultats complémentaires

Khaled Belahcene/Sylvain Lagrue

sylvain.lagrue@hds.utc.fr

I. Rappels

Résumé des épisodes précédents

En théorie du choix social, on s'intéresse à la prise de décision collective d'agents :

- Collaboratifs (ils se soumettent à la décision collective)
- Sincères (ils expriment leurs véritables préférences)

On est alors amené à définir une *fonction de bien-être social* (SWF, qui agrège les préférences de chaque votant en un ordre collectif), ou de *choix social* (SCF, qui identifie un ou plusieurs vainqueur(s)).

On peut aborder ce problème sous divers angles :

- Algorithmique
- Axiomatique
- Recherche de consensus
- Recherche d'une vérité latente

I. Rappels

Approche algorithmique

C'est l'approche principalement suivie dans le cours. Nous avons défini, puis mis en œuvre, diverses procédures de vote, et nous avons examiné certaines de leurs propriétés (par exemple *monotonie*, *IIA*, *Condorcet*, etc.).

Approche axiomatique

On part d'un cahier des charges recensant les propriétés désirables, vues comme autant de contraintes, et en cherchant à identifier l'ensemble des procédures les satisfaisant toutes. Ici, procédure = SCF (on s'intéresse à l'ensemble des vainqueurs)

I. Rappels

Exemple d'axiomes pour les SCF décisives

- N l'ensemble des votants, A l'ensemble des candidats, v la règle de vote (SCF décisive)
- Pour un profil $\mathbf{P}$ et deux alternatives $x \neq y$, on note $N^{\mathbf{P}}_{x \succ y}$ l'ensemble des votants préférant x à y dans $\mathbf{P}$
- *Unanimité :* pour tout $x, y, \mathbf{P}$

$$N^{\mathbf{P}}_{x \succ y} = N \Rightarrow y \neq v(\mathbf{P})$$

- *Indépendance :* pour tout $x, y, \mathbf{P}, \mathbf{P}'$

$$v(\mathbf{P}) = x \text{ et } N^{\mathbf{P}}_{x \succ y} = N^{\mathbf{P}'}_{x \succ y} \Rightarrow v(\mathbf{P}') \neq y$$

- *Monotonie (forte)* : pour tout $x, \mathbf{P}, \mathbf{P}'$,

$$v(\mathbf{P}) = x \text{ et } \forall y \in A \setminus x \; N^{\mathbf{P}}_{x \succ y} \subseteq N^{\mathbf{P}'}_{x \succ y} \Rightarrow v(\mathbf{P}') = x$$

- ...

I. Rappels

Quelques résultats

> *Les procédures qui satisfont Domaine universel, IIA et Unanimité sont les dictatures (Arrow, 1951)*

> *Pour 2 candidats, le vote majoritaire est la seule procédure satisfaisant les propriétés de Domaine universel, Anonymat, Neutralité et Monotonie (May, 1952)*

> *Les méthodes de vote par score sont exactement celles qui satisfont Neutralité, Anonymat, Renforcement et Séparabilité (Young, 1974)*

> *Pour au moins 3 candidats, les méthodes Décisives, Surjectives et Fortement monotones sont les dictatures (Muller and Satterthwaite, 1977)*

II. Théorie du vote et recherche de consensus

Recherche de consensus

- *Rappel :* les procédures de Dodgson et de Kemeny-Young élisent le vainqueur le plus "proche" d'un vainqueur de Condorcet
- *Principe général:* définir des procédures en définissant
  - Une classe de *profils consensuels*, pour lequel le vainqueur est clair
  - Une façon de *mesurer la distance* entre un rangement et un profil de vote
  - Le résultat de la SWF comme l'un des *rangements minimisant la distance avec les suffrages*, ou de la SCF comme étant l'*élément maximal* de ce rangement

II. Théorie du vote et recherche de consensus

Profils consensuels

- *Vainqueur de Condorcet :* il y a un vainqueur de Condorcet
- *Vainqueur majoritaire :* il y a un candidat classé premier par une majorité absolue de votants
- *Vainqueur unanime :* il y a un candidat classé premier par la totalité des votants
- *Classement unanime :* tous les votants expriment le même rangement

II. Théorie du vote et recherche de consensus

Distance entre un rangement et un profil

On définit souvent la distance entre un rangement $\ell \in \mathcal{L}$ et un profil $\mathcal{P} \in \mathcal{L}^N$

- En définissant une *distance inter-rangements* $d : \mathcal{L} \times \mathcal{L} \to \mathbb{R}^+$ vérifiant les axiomes de symétrie, de séparation et l'inégalité triangulaire,
- En *agrégeant* ces distances à l'aide d'une fonction d'agrégation $\varphi \in \mathbb{R}^N \to \mathbb{R}$

En pratique :

- Il y a deux distances inter-rangements canoniques : la *distance d'édition* (la fonction τ de Kendall compte le nombre de paires discordantes entre les deux rangements) et la *distance discrète* (ou *drastique*, qui vaut 0 en cas d'égalité, 1 sinon)
- L'agrégation est le plus souvent réalisée à l'aide d'une *somme* (classement *médian*, c'est souvent facile, et ça a de bonnes propriétés pour l'approche suivante), mais on pourrait utiliser la *somme des carrés* (classement *moyen*) ou le *maximum* (classement *prudent*)

II. Théorie du vote et recherche de consensus

Exemples (*distance rationalizability*)

Par définition

- Règle de Dodgson = Vainqueur de Condorcet + distance d'édition
- Règle de Kemeny = Rangement unanime + distance d'édition

Et les autres ?

- Règle de Borda = Vainqueur unanime + distance d'édition

(Farkas and Nitzan, 1979)

- Règle majoritaire = Vainqueur unanime + distance discrète

(Nitzan, 1981)

Loading [MathJax]/extensions/MathZoom.js

III. Théorie du vote et recherche d'une vérité latente

Conitzer, V. and Sandholm, T. Common Voting Rules as Maximum Likelihood Estimators.

Hypothèses

- Il existe un rangement « véritable » ou « correct » latent
- Les préférences exprimées correspondent à des observations bruitées de ce rangement
- Le bruit est I.I.D. (indépendant et identiquement distribué)

III. Théorie du vote et recherche d'une vérité latente

Estimateur du maximum de vraisemblance (EMV)

On cherche alors à retrouver le « vrai » rangement, qui nous est inconnu, à partir de ces observations bruitées.

Étant donné un *vrai rangement*, et un *modèle du bruit*, on peut calculer le rangement pour lequel le profil de préférences observé est *le plus probable*.

Théorèmes

- Pour le modèle de *bruit de Condorcet* (constant), l'EMV est la règle de *Kemeny*
- Toute règle *SCF* fondée sur un *score de position* est un EMV pour *un certain modèle de bruit* (en fixant la probabilité d'observer un profil classant le vrai vainqueur en position j proportionnelle à 2^{s_j})
- Toute règle *SWF* fondée sur un score de position est aussi un EMV (avec un modèle de bruit différent)
- *Copeland* n'est *pas un EMV*, que ce soit en tant que SWF ou que SCF
- *STV* est un EMV comme SWF, mais pas comme SCF

III. Théorie du vote et recherche d'une vérité latente

Utilisation de l'information

Sans être un EMV, on peut vouloir exiger d'une SWF/SCF qu'elle soit en mesure d'identifier la « vérité » en disposant d'un *échantillon infini*… et se préoccuper de *vitesse de convergence*

III. Théorie du vote et recherche d'une vérité latente

Elle a l'air drôlement bien, cette règle de Kemeny

- Définie comme un consensus
- EMV pour le bruit de Condorcet
- Respecte le critère de Condorcet
- **est NP-difficile à calculer !** :'(

Loading [MathJax]/extensions/MathZoom.js

IV. Théorie du vote et agents stratégiques

Votants stratégiques

Si le but de la procédure de vote n'est pas de manifester une réalité latente, mais d'établir un consensus entre les agents, un agent peut avoir intérêt à mentir pour faire prévaloir ses intérêts.

IV. Théorie du vote et agents stratégiques

Voter sincèrement ?

Votants	1	2	3	4	5
Rang 1	a	b	e	e	?
Rang 2	b	a	c	c	?
Rang 3	c	d	b	b	?
Rang 4	d	e	a	a	?
Rang 5	e	c	d	d	?

On suppose qu'en cas d'égalité, $a \succ_{tb} b \succ_{tb} c \succ_{tb} d \succ_{tb} e$ et supposons que $c \succ_5 d \succ_5 b \succ_5 a \succ_5 e$.

5 a-t-il intérêt à voter conformément à ses préférences si le vainqueur est désigné :

- À la majorité simple ? à la majorité à deux tours ? au décompte de Borda ?
- Par la méthode de Copeland ? par la méthode du vote unique transférable (STV) ?

IV. Théorie du vote et agents stratégiques

Formalisation

On s'intéresse aux fonctions de choix social décisives, c.-à-d. qui associent un vainqueur unique à chaque profil de vote

Manipulabilité

Une règle de vote est *manipulable* si un électeur peut soumettre un profil de préférence fallacieux et obtenir un résultat plus favorable que s'il avait révélé ses préférences

Formellement, v est *manipulable* s'il existe un profil $\succ$, un électeur i et un bulletin non sincère $\succ_i'$ tel que $v(\succ_{-i} \cup \succ_i') \succ_i v(\succ)$.

IV. Théorie du vote et agents stratégiques

Existe-t-il des règles de vote non manipulables ?

Certes, oui…

- **Dictature** : le vainqueur est celui choisi par l'agent i^*

- **Règle constante** : le vainqueur est a^*, indépendamment des suffrages

Loading [MathJax]/extensions/MathZoom.js

IV. Théorie du vote et agents stratégiques

Existe-t-il des règles de vote non manipulables ?

Théorème (Gibbard-Satterthwaite)

> *Pour plus de 3 candidats, toute règle de vote non-manipulable et surjective est une dictature.*

Rappel : pour plus de 3 candidats, toute règle de vote unanime et indépendante est une dictature (Arrow)

IV. Théorie du vote et agents stratégiques

C'est grave ?

Oui !

- Mentir, c'est pas bien
- Les électeurs ne devraient pas avoir à consacrer du temps et des ressources pour deviner ce que les autres électeurs vont faire, et comment y répondre au mieux
- Si tout le monde manipule, le système n'est pas forcément stable (*problème du vote itéré*)
- Si tout le monde manipule et se trompe, le résultat final risque de ne plus correspondre aux souhaits du corps électoral

IV. Théorie du vote et agents stratégiques

Mitigation

- Le théorème ne s'applique qu'aux *règles décisives*. En pratique, elles ne le sont pas !

> *(Duggan and Schwartz, 2000) ont prouvé un théorème similaire dans le cadre des SCF indécisives*

- *restriction de domaine :* toute règle de vote est manipulable à condition d'accepter n'importe quel profil. Des restrictions "raisonnables" peuvent empêcher ou réduire la possibilité de manipulation
    - Lorsque les profils sont à *pic simple*, la règle du vote majoritaire n'est pas manipulable

IV. Théorie du vote et agents stratégiques

- Le théorème de Gibbard-Satterthwaite garantit l'existence de manipulations dans certains cas. Sont-ils *"fréquents"* ?

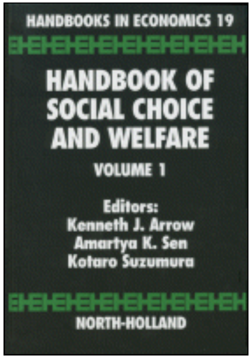
  > *Malheureusement oui, d'autant plus si on considère les manipulations coalitionnelles*

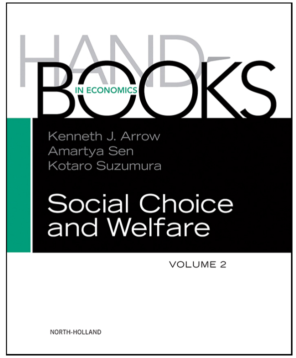
- Le théorème de Gibbard-Satterthwaite n'est pas constructif : il ne dit pas comment calculer une manipulation. Peut-être est-ce difficile ?

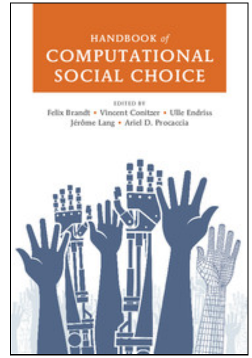
  > *Pour un électeur isolé, c'est facile (P) pour les PSR (positional scoring rules), difficile (NP-complet) pour certaines règles (STV notamment). Pour une coalition en vote pondéré, c'est difficile même pour Borda !*

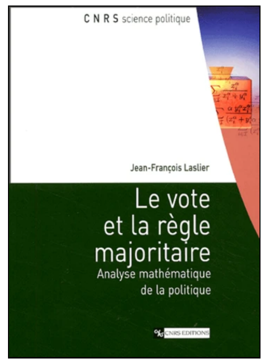
- Le théorème de Gibbard-Satterthwaite suppose une information parfaite des électeurs
- ...

Bibliographie









- *Handbook of Social Choice and Welfare (vol. 1)* : Kenneth J. Arrow, Amartya Sen, Kotaro Suzumura (2002)
- *Handbook of Social Choice and Welfare (vol. 2)* : Kenneth J. Arrow, Amartya Sen, Kotaro Suzumura (2011)
- *Handbook of Computationnal Social Choice* : Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, Ariel D. Procaccia (2016)
- : Steven Brams, Peter C. Fishburn (2007)

Pour s'avancer...

Exercice de TP

> *Définition : en présence d'information incomplète sur les préférences (par exemple, lorsque l'on connait les votes de tous les électeurs sauf un), on dit qu'un candidat est :*
>
> - *vainqueur nécessaire s'il gagne l'élection pour toutes les manières possibles de compléter les préférences,*
> - *vainqueur possible s'il gagne l'élection pour au moins une manière possible de compléter les préférences,*
> - *vainqueur impossible s'il ne gagne l'élection dans aucune configuration possible.*

Exercice

1. En faisant appel aux fonctions déjà réalisées qui implémentent différentes règles de vote, programmer une fonction `possibleWinners(P Profile) map[candidate] []candidate` qui, étant donné le profil de vote de tous les électeurs sauf un, détermine les vainqueurs possibles de l'élection, ainsi qu'une des complétions correspondantes. On procèdera de manière *brutale*, en énumérant tous les profils de vote pour l'électeur manquant.
2. A l'aide de cette fonction, programmer `isPossibleWinner(P Profile, c candidate) bool` et `isNecessaryWinner(P Profile, c candidate) bool`.
3. A l'aide de `possibleWinners`, programmer la fonction `bestResponse(P profile, pref []candidate) (winner candidate, ballot [] candidate)` qui, étant donné un profil incomplet P et les préférences de l'électeur manquant, renvoie la meilleure stratégie de vote de ce candidat (en cas d'ex aequo, on renverra un des bulletins les plus proches des préférences sincères du candidat au sens de la distance d'édition).