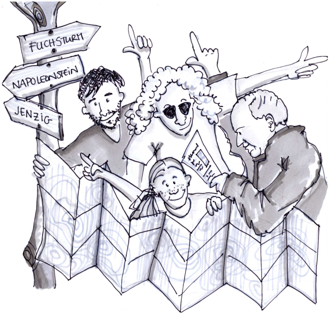
A Multivariate Complexity Analysis of Voting Problems

Nadja Betzler
Friedrich-Schiller-Universität Jena



GI-Dissertationspreis 2010 – Kolloquium, Dagstuhl

Wahlsituationen

Wahl: Prozess einer gemeinschaftlichen Entscheidungsfindung

- Politische Wahlen / Bürgerentscheide
- Preisverleihungen
- Bewerberauswahl
- Multiagentensysteme / Metasuchmaschinen
- . . .


Wahl

Wähler entscheiden über eine Menge von Kandidaten.

Wahl

Stimmabgaben über einer Menge von **Kandidaten**.
Eine Stimmabgabe (**Ranking**) ist ein totale Ordnung.

Auswertung erfolgt anhand eines **Wahlsystems**.

Beispiel:

Johannes:	Rom > Paris > London > Amsterdam
Christian:	London > Paris > Amsterdam > Rom
Sepp:	Rom > Paris > Amsterdam > London

Untersuchte Probleme

Teil I: Gewinnerbestimmung

Wie werte ich am schnellsten aus, wer gewinnt?

Teil II: Möglicher Gewinner

Unvollständige Information: Kann mein Favorit noch gewinnen?

Teil III: Wahlkontrolle

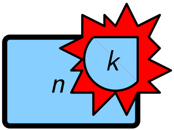
Kann ich meinen Favoriten durch Löschen oder Hinzufügen von Kandidaten zum Gewinner machen?

NP-harte Probleme für viele Wahlsysteme

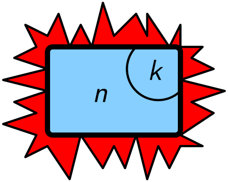
Methodik: Parametrisierte Algorithmik

Betrachte: NP-hartes Problem, Eingabegröße n und Parameter k

Grundidee: Beschränke die kombinatorische Explosion durch k



anstatt



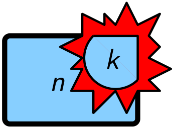
Definition

Ein Problem ist **fixed-parameter tractable** bezüglich k, wenn jede Instanz in $f(k) \cdot \text{poly}(n)$ Zeit entschieden werden kann.

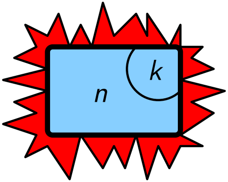
Methodik: Parametrisierte Algorithmik

Betrachte: NP-hartes Problem, Eingabegröße n und Parameter k

Grundidee: Beschränke die kombinatorische Explosion durch k



anstatt



Definition

Ein Problem ist **fixed-parameter tractable** bezüglich k, wenn jede Instanz in $f(k) \cdot \text{poly}(n)$ Zeit entschieden werden kann.

$$\text{FPT} \subseteq \overbrace{\text{W[1]} \subseteq \text{W[2]} \subseteq \ldots \subseteq \text{W[P]}}^{\text{Wahrscheinlich nicht fixed-parameter tractable}}$$

Untersuchte Probleme

Teil I: Gewinnerbestimmung

Wie werte ich am schnellsten aus, wer gewinnt?

Teil II: Möglicher Gewinner

Unvollständige Information: Kann mein Favorit noch gewinnen?

Teil III: Wahlkontrolle

Kann ich meinen Favoriten durch Löschen oder Hinzufügen von Kandidaten zum Gewinner machen?

Parameter: Anzahl der Wähler, Anzahl der Kandidaten, **problem-spezifische Parameter**

Kemeny Ranking

Kemeny Ranking

Gegeben: Eine Multimenge von Rankings.
Ziel: Suche ein Konsensranking, das die Summe der Abstände zu den Eingaberankings minimiert.

Abstand zwischen 2 Rankings (**KT-Distanz**): # der Inversionen

r_1 :	$a > b > c$	KT-dist$(r_{con}, r_1) = 0$
r_2 :	$a > c > b$	KT-dist$(r_{con}, r_2) = 1$ wegen $\{b, c\}$
r_3 :	$b > c > a$	KT-dist$(r_{con}, r_3) = 2$ wegen $\{a, b\}$ und $\{a, c\}$
r_{con} :	$a > b > c$	Kemeny score: $0 + 1 + 2 = 3$

Kemeny Ranking

Kemeny Ranking

Gegeben: Eine Multimenge von Rankings.
Ziel: Suche ein Konsensranking, das die Summe der Abstände zu den Eingaberankings minimiert.

Abstand zwischen 2 Rankings (**KT-Distanz**): # der Inversionen

r_1 :	$a > b > c$	$\text{KT-dist}(r_{con}, r_1) = 0$
r_2 :	$a > c > b$	$\text{KT-dist}(r_{con}, r_2) = 1$ wegen $\{b, c\}$
r_3 :	$b > c > a$	$\text{KT-dist}(r_{con}, r_3) = 2$ wegen $\{a, b\}$ und $\{a, c\}$
r_{con} :	$a > b > c$	Kemeny score: $0 + 1 + 2 = 3$

Kemeny Score ist NP-vollständig (für 4 Rankings).

[Dwork, Kumar, Naor & Sivakumar, WWW 2001]

Motivation Kemeny

Anwendungsgebiete:

- Metasuchmaschinen, Erkennung von Spam [Dwork et al., WWW 2001]
- Datenbanken [Fagin et al., SIGMOD, 2003]
- Bioinformatik [Jackson et al., IEEE/ACM Transactions on Computational Biology and Bioinformatics 2008]

Algorithmen:

- Randomisierte Factor-11/7 Approximation [Ailon et al., Journal of the ACM 2008]
- PTAS [Kenyon-Mathieu & Schudy, STOC 2007]
- Heuristiken; Greedy, Branch&Bound [Davenport & Kalagnanam, AAAI 2004],[Conitzer et al. AAAI 2006], [Schalekamp & van Zuylen, ALENEX 2009]

Strukturelle Parameter

Untersuchte Fragestellung

Kann man algorithmisch ausnutzen, dass sich die Rankings „ähnlich" sind?

Parameter:

- maximale und durchschnittliche KT-Distanz
- maximaler und durchschnittlicher „Range" der Kandidaten

$$\begin{array}{|ccccccccccc|}\hline a & > & c & > & b & > & e & > & d & > & f \\ b & > & a & > & c & > & d & > & e & > & f \\ b & > & c & > & a & > & e & > & f & > & d \\ \hline\end{array}$$

$$\begin{array}{|ccccccccccc|}\hline b & > & c & > & d & > & e & > & f & > & a \\ a & > & b & > & c & > & d & > & e & > & f \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ a & > & b & > & c & > & d & > & e & > & f \\ \hline\end{array}$$

Parametrisierte Komplexität Kemeny Ranking

Range der Kandidaten:

- Durchschnittlicher Range r_a: NP-hart für jedes $r_a \geq 2$
- Maximaler Range r: 32^r (dynamisches Programmieren)

Durchschnittliche KT-Distanz d der Rankings:

- 16^d (dynamisches Programmieren)
- 5.83^d [Simjour, IWPEC 2009]
- $2^{O(\sqrt{d})}$ [Karpinski & Schudy, ISAAC 2010]

Ergänzender Ansatz: Polynomielle Datenreduktion, so dass die verbleibende Instanz höchstens $11 \cdot d$ Kandidaten hat.

$\geq_{3/4}$-Mehrheiten

$a \geq_{3/4} b$

gdw. Kandidat a in mindestens 3/4 der Rankings vor b steht.

Kandidat c ist **nondirty**, wenn für jeden anderen Kandidaten b entweder

$$c \geq_{3/4} b \text{ oder } b \geq_{3/4} c.$$

$$\boxed{a_1} > a_2 > a_3 > \boxed{c} > b_1 > b_2$$
$$a_3 > a_2 > \boxed{c} > \boxed{a_1} > b_2 > b_1$$
$$\boxed{a_1} > \boxed{c} > a_2 > b_2 > b_1 > a_3$$
$$a_2 > a_3 > \boxed{a_1} > b_1 > b_2 > \boxed{c}$$

$\geq_{3/4}$-Mehrheiten

a $\geq_{3/4}$ b

gdw. Kandidat a in mindestens $3/4$ der Rankings vor b steht.

Kandidat c ist **nondirty**, wenn für jeden anderen Kandidaten b entweder

$$c \geq_{3/4} b \text{ oder } b \geq_{3/4} c.$$

$$a_1 > a_2 > a_3 > c > b_1 > b_2$$
$$a_3 > a_2 > c > a_1 > b_2 > b_1$$
$$a_1 > c > a_2 > b_2 > b_1 > a_3$$
$$a_2 > a_3 > a_1 > b_1 > b_2 > c$$

$a_i \geq_{3/4} c$ und $c \geq_{3/4} b_i$

$\Rightarrow$

Kemeny Ranking:

$\{a_1, a_2, a_3\} > c > \{b_1, b_2\}$

Datenreduktionsregel

Wenn c nondirty, lösche c und teile Instanz in zwei Teilinstanzen T_1 und T_2, so dass T_1 genau alle Kandidaten mit $a \geq_{3/4} c$ enthält.

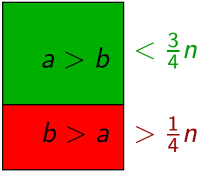
Analyse

Nach Datenreduktion: nur noch „dirty" Kandidaten.

Idee

Wenn die durchschnittliche KT-Distanz klein ist, dann gibt es nur beschränkt viele „dirty pairs".

Für n Rankings:



Beitrag zur durchschnittlichen KT-Distanz pro dirty pair $\{a, b\}$

$$> \frac{3}{4}n \cdot \frac{1}{4}n \cdot \frac{2}{n(n-1)} > \frac{6}{16}$$

Theorem

Nach Anwendung der Reduktionsregel enthält eine Kemeny-Instanz mit durchschnittlicher KT-Distanz d höchstens $11 \cdot d$ Kandidaten.

Experimente: Metasuchmaschinen

4 Rankings: Google, Lycos, MSN Live Search, und Yahoo!

jeweils top 1000; Kandidaten, die in allen 4 Rankings vorkommen

Suchwort	#Kand.	Zeit [s]	Struktur reduzierte Instanz gelöst/ungelöst		
affirmative action	127	0.41	[27]	> **41** >	[59]
alcoholism	115	0.21	[115]		
architecture	122	0.47	[36]	> **12** > [30] > **17** >	[27]
blues	112	0.16	[74]	> **9** >	[29]
cheese	142	0.39	[94]	> **6** >	[42]
classical guitar	115	1.12	[6]	> **7** > [50] > **35** >	[17]
Death+Valley	110	0.25	[15]	> **7** > [30] > **8** >	[50]
field hockey	102	0.21	[37]	> **26** > [20] > **4** >	[15]
gardening	106	0.19	[54]	> **20** > [2] > **9** > [8] > **4** >	[9]
HIV	115	0.26	[62]	> **5** > [7] > **20** >	[21]
lyme disease	153	2.61	[25]	> **97** >	[31]
mutual funds	128	3.33	[9]	> **45** > [9] > **5** > [1] > **49** >	[10]
rock climbing	102	0.12	[102]		
Shakespeare	163	0.68	[100]	> **10** > [25] > **6** >	[22]
telecommuting	131	2.28	[9]	> **109** >	[13]

Parametrisierte Komplexität von Kemeny

	Kemeny Score	with ties	incomplete votes
# votes n	NP-h for $n = 4$ (♣)	NP-h for $n = 4$ (♣)	NP-h for $n = 4$ (♣)
# candidates m	2^m	2^m	2^m
Kemeny score k	$2^{O(\sqrt{k})}$ (♢)	1.76^k	$k! \cdot 4^k$
max. range r_m	32^{r_m}	$(3r_m + 1)! \cdot 2^{3r_m+1}$	—
avg. range r_a	NP-h for $r_a \geq 2$	NP-h for $r_a \geq 2$	—
max. KT-dist d_m	$2^{O(\sqrt{d_m})}$ (♢)	$(6d_m + 2)! \cdot 2^{6d_m+2}$	NP-h for $d_m = 0$
avg. KT-dist d_a	$2^{O(\sqrt{d_a})}$ (♢)	$2^{O(d_a^2)}$	NP-h for $d_a = 0$
above guarantee	FPT (♠)	?	?

(♣) [Dwork et al. WWW 2001]
(♢) [Karpinski & Schudy, ISAAC 2010]
(♠) [Mahajan et al., JCSS 2009]
Ergebnisse dieser Arbeit

Verbesserte Ergebnisse

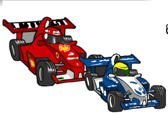
Untersuchte Probleme

Teil I: Gewinnerbestimmung

Wie werte ich am schnellsten aus, wer gewinnt?

Teil II: Möglicher Gewinner

Unvollständige Information: Kann mein Favorit noch gewinnen?



Teil III: Wahlkontrolle

Kann ich meinen Favoriten durch Löschen oder Hinzufügen von Kandidaten zum Gewinner machen?

Unvollständige Information

Bisher: Man hat die vollständigen Rankings von allen Wählern

- Einige Wähler haben noch nicht gewählt.
- Es kommen neue Kandidaten hinzu.
- Ein Wähler kann einige Kandidaten nicht vergleichen.

$\Rightarrow$ Information liegt in Form von partiellen Rankings vor.

Möglicher Gewinner

Eingabe: Ein Wahlsystem, eine Menge von Kandidaten, eine Multimenge V von partiellen Rankings und ein favorisierter Kandidat c.
Frage: Kann V so ergänzt werden, dass c gewinnt?

Scoring-Regeln

Scoring-Regel

Für m Kandidaten enthält ein **Scoring-Vektor** $(\alpha_1, \alpha_2, \ldots, \alpha_m)$ eine Punktzahl für jede Position.
Gewinner: Kandidat mit höchster Gesamtpunktzahl.

Beispiele:

- k-Approval $(1, \ldots, 1, 0, \ldots, 0, 0, 0)$ mit k Einsen
  - Plurality: $(1, 0, \ldots, 0)$
  - Veto: $(1, \ldots, 1, 0)$
- Borda: $(m-1, m-2, \ldots, 0)$
- Formel 1: $(25, 18, 15, 12, 10, 8, 6, 4, 2, 1, 0, \ldots, 0)$
- Jenaer Bürgerbefragung: $(3, 2, 1, 0, \ldots, 0)$

Ergebnisse I

Dichotomie

MÖGLICHER GEWINNER Problem:

- Für Plurality und Veto in Polynomialzeit entscheidbar
- Für alle anderen Scoring-Regeln NP-vollständig

- vorher: NP-Härte-Resultate für eine Klasse von Scoring-Regeln [Xia & Conitzer, AAAI 2008]
- Beispiele für neue Ergebnisse dieser Arbeit:
  - $(1, 1, 0, \ldots, 0)$
  - $(3, 2, 1, 0, \ldots, 0)$
  - $(3, 1, \ldots, 1, 0)$
- Nachfolgearbeit: $(2, 1, \ldots, 1, 0)$ [Baumeister & Rothe, ECAI 2010]

Ergebnisse II: Parameterisierte Komplexität

Einzelne Parameter:

- Anzahl Kandidaten: FPT für alle Scoring-Regeln
- Anzahl Wähler:
  - Borda: NP-hart für drei partielle Rankings (und drei Rankings)
  - k-Approval: NP-hart für zwei partielle Rankings
- Parameter, die den Grad der Unvollständigkeit messen

Kombinierter Parameter

Ist MÖGLICHER GEWINNER für k-Approval effizient lösbar, wenn k klein ist **und** es nur wenige Wähler gibt?

Ergebnisse II: Parameterisierte Komplexität

Einzelne Parameter:

- Anzahl Kandidaten: FPT für alle Scoring-Regeln
- Anzahl Wähler:
  - Borda: NP-hart für drei partielle Rankings (und drei Rankings)
  - k-Approval: NP-hart für zwei partielle Rankings
- Parameter, die den Grad der Unvollständigkeit messen

Kombinierter Parameter

Ist MÖGLICHER GEWINNER für k-Approval effizient lösbar, wenn k klein ist **und** es nur wenige Wähler gibt?

Ja, allerdings gibt es keinen polynomiellen Problemkern.

Untersuchte Probleme

Teil I: Gewinnerbestimmung

Wie werte ich am schnellsten aus, wer gewinnt?

Teil II: Möglicher Gewinner

Unvollständige Information: Kann mein Favorit noch gewinnen?

Teil III: Wahlkontrolle

Kann ich meinen Favoriten durch Löschen oder Hinzufügen von Kandidaten zum Gewinner machen?

Ausgewählte Publikationen

- N. Betzler, J. Guo, and R. Niedermeier. **Parameterized Complexity of Dodgson and Young Elections.** *Information and Computation*, 2010
- N. Betzler, M. R. Fellows, J. Guo, R. Niedermeier, and F. A. Rosamond. **Fixed-Parameter Algorithms for Kemeny Rankings.** *Theoretical Computer Science*, 2009
- N. Betzler and B. Dorn. **Towards a Dichotomy of Finding Possible Winners in Elections Based on Scoring Rules.** *Journal of Computer and System Sciences*, 2010
- N. Betzler, S. Hemmann, and R. Niedermeier. **A Multivariate Complexity Analysis of Determining Possible Winners Given Incomplete Votes.** *Twenty-first International Joint Conference on Artificial Intelligence (IJCAI-09)*
- N. Betzler. **On Problem Kernels for Possible Winner Determination Under the k-Approval Protocol.** *Algorithmica*, accepted due to minor revision.
- N. Betzler and J. Uhlmann. **Parameterized Complexity of Candidate Control in Elections and Related Digraph Problems.** *Theoretical Computer Science*, 2009

Vielen Dank!