

KIWIEL



POLSKA AKADEMIA NAUK
Instytut Badań Systemowych

WSPOMAGANIE DECYZJI

SYSTEMY EKSPERCKIE

pod redakcją

Romana Kulikowskiego i Lucyny Bogdan

Warszawa 1995

WSPOMAGANIE DECYZJI

SYSTEMY EKSPERCKIE

pod redakcją

Romana Kulikowskiego i Lucyny Bogdan

Warszawa 1995

Wydano z wykorzystaniem dotacji
KOMITETU BADAŃ NAUKOWYCH

Materiały konferencji: "Analiza Decyzyjna, Systemy Ekspertckie, Zastosowania Systemów Komputerowych",
Warszawa, 25-27 maja 1994r.

Komitet Programowy Konferencji:

Andrzej Ameljańczyk, Zdzisław Bubnicki, Wiesław Grudzewski, Olgierd Hryniewicz, Janusz Kacprzyk, Lech Kruś, Roman Kulikowski (przewodniczący), Kazimierz Mańczak, Ireneusz Nykowski, Zdzisław Pawlak, Roman Słowiński, Andrzej Straszak, Andrzej Weryński, Andrzej Wierzbicki.

Wykonano z oryginałów tekstowych dostarczonych przez autorów

© Instytut Badań Systemowych PAN, Warszawa 1995

ISBN 83-85847-85-5

PROGNOZOWANIE WYNIKÓW WYBORÓW PARLAMENTARNYCH KONCEPCJA WYKORZYSTANIA ZBIORÓW PRZYBLIŻONYCH

Jerzy Hołubiec i Janusz Kacprzyk
Instytut Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszawa
E-mail: kacprzyk@ibspan.waw.pl

I. WPROWADZENIE

W ostatnich kilku latach w wielu krajach świata, również w Polsce, wyniki wyborów parlamentarnych zaskoczyły zarówno specjalistów (komentatorów politycznych, polityków, politologów itp.) jak i zwykłych obywateli. Zwyciężyły bowiem w nich partie, na które niewielu stawiało przed wyborami, a przegrały te, które wydawały się zdecydowanymi faworytami.

W tym kontekście narzucają się więc od razu liczne pytania, jak np. dlaczego tak się stało, czy można to było jakoś przewidzieć, czy znajomość wyników z poprzednich wyborów ma jakieś znaczenie itp. W naszym przypadku sprawa wyjaśnienia przyczyn takich czy innych wyników wyborów wychodzi oczywiście poza zakres naszych zainteresowań, bo dotyczy np. różnych aspektów zachowań wyborców i ma więcej wspólnego np. z naukami społecznymi niż z analizą systemową czy teorią głosowań (por. Hołubiec i Mercik, 1992).

Z punktu widzenia naszych zainteresowań ciekawa jest natomiast sprawa ewentualnego wykorzystania znajomości pewnych dostępnych danych dotyczących np. programów partii, "image'u" przywódców, itp. w celu uzyskania wskazówek (może nawet przewidzenia?) wyników wyborów.

Do tego celu można zastosować wiele metod, a jedną z najpopularniejszych jest metoda z użyciem tzw. indeksów siły np. Shapleya-Shubika czy Banzhafa (por. Hołubiec i Mercik, 1992; Gambarelli, 1994). Metodę tę zastosowano (Hołubiec, 1994) biorąc najpierw, jako punkt wyjścia do analizy, wyniki wyborów do Sejmu RP w 1991 r.

Otóż, w ich rezultacie do Sejmu weszli przedstawiciele 29 partii politycznych, co spowodowało duże rozdrobnienie miejsc i związane z tym rozproszenie głosów, a co za tym idzie, duże trudności w sprawnym funkcjonowaniu Sejmu, m.in. w zakresie prac legislacyjnych.

Aby uniknąć takich komplikacji, wprowadzono do regulaminu wyborów w 1993 r. próg 5% głosów do wejścia partii do Sejmu. Biorąc to ograniczenie pod uwagę i podział głosów na poszczególne partie polityczne w poprzednich wyborach (1991 r.), przy dokonaniu korekt w celu uwzględnienia zmian na scenie politycznej kraju, jakie nastąpiły w międzyczasie, wyznaczono partie, które w wyborach 1993 r. powinny ten próg 5% przekroczyć. Obliczenia, z użyciem technik opartych na indeksach siły, dały tu w wyniku następujących 9 partii: UD, SLD, ZChN, PC, PSL, KPN, KLD, PL i NSZZ "Solidarność". Określono także dla tych partii hipotetyczne wartości indeksów siły. Ponadto, rozpatrzono różne koalicje, jakie w takim układzie mogły powstać w celu sformowania rządu, dysponującego wymaganą do sprawowania władzy większością w Sejmie, zapewniającą stabilność władzy.

Nietrudno zauważyć, że otrzymane powyżej wyniki wyborów 1993 r. otrzymane na podstawie wyników poprzednich z 1991 r. (wraz z odpowiednimi korektami) znacznie się różnią od wyników rzeczywistych z 1993 r. Do Sejmu weszło wtedy bowiem 7 partii: SLD, PSL, UD, UP, KPN, BBWR (nowe ugrupowanie polityczne) i mniejszość niemiecka (nie obowiązywał dla niej próg 5%).

Tak duże rozbieżności między rzeczywistością a prognozą są spowodowane wieloma przyczynami, jak np. "młodość" polskiej demokracji, do czego "nie pasuje" zaawansowane metody formalne opracowane w ustabilizowanych demokracjach zachodnich, niedokładnie sprecyzowane i niezbyt zrozumiałe dla społeczeństwa cele, programy, metody działania poszczególnych partii itp.

Należałoby być może zastosować inne, jakościowo różne podejście do opracowania takiej prognozy. Koncepcję takiego nowego podejścia zarysowujemy w niniejszej pracy. Ogólnie biorąc ma ono związek z tzw. *inteligentną analizą danych*. Otóż, naszym punktem wyjścia są tu opisy poszczególnych partii przez spełnienie i/lub niespełnienie pewnego zestawu kryteriów (jakościowych i ilościowych) dot. programów, cech partii i ich przywódców itp. Takie dane są następnie poddawane "inteligentnej" analizie w celu wykrycia tych aspektów (kryteriów), które rzeczywiście różnicują poszczególne partie. Do tego celu stosujemy pewne elementy teorii *zbiorów przybliżonych* Pawlaka (1982, 1991).

Wykorzystanie wyników powyższej analizy odbywa się potem w następujący sposób. Znajac aspekty, które różnicują poszczególne partie polityczne oraz (np. z badań socjologicznych, opinii ekspertów itp.) preferencje elektoratu dotyczące tych aspektów, dokonujemy analizy partii politycznych z punktu widzenia najbardziej preferowanych (istotnych) aspektów i partie wypadające najlepiej z tego punktu widzenia wchodzi do prognozy wyników wyborów. Oczywiście, można tej analizie dokonywać na poziomie jakościowym, ale można ją wzbogacić o aspekty ilościowe, np. wprowadzając intensywność powyższych preferencji.

Niniejsza praca, która stanowić będzie punkt wyjścia do formalnego przedstawienia powyższej metodologii, jest poświęcona pierwszemu aspektowi, tzn.

inteligentnej analizie danych wyborczych z użyciem zbiorów przybliżonych Pawlaka (1982, 1991).

2. ZBIORY PRZYBLIŻONE I ICH PODSTAWOWE WŁAŚCIWOŚCI

Pojęcie *zbioru przybliżonego* (and. rough set) wprowadził Pawlak (1982) (por. także Pawlak, 1991). Najogólniej biorąc zbiór przybliżony jest środkiem formalnym służącym do reprezentacji nieprecyzyjności znaczeń rozumianej tak, że klasyfikowane pojęcie jest samo w sobie nieprecyzyjne, ale może być aproksymowane pojęciami precyzyjnymi (np. utożsamianymi ze zbiorami konwencjonalnymi). Zauważmy, że jest to sytuacja inna niż w przypadku rozmytości, gdzie nieprecyzyjność ma charakter częściowej przynależności do zbiorów.

Założmy, że U jest obszarem rozważań, który można podzielić na wiele sposobów. Jako tzw. *bazę wiedzy* określa się parę $K=(U, R)$ składająca się z U i rodziny R podziałów lub relacji równoważności w U . Założmy, że $\emptyset \neq P \subseteq R$. Przecięcie wszystkich relacji równoważności (lub podziałów) w P jest *relacją nieróżnorodności* w P ; oznacza się ją jako $IND(P)$ i nazywa *wiedzą P -podstawową* o U w bazie wiedzy K . Określa ona oczywiście wszystko to, co można powiedzieć o elementach U , jeśli tylko P jest dostępna. W szczególności, nie można sklasyfikować elementów U "głębiej" niż do poziomu $IND(P)$.

Klasy równoważności $IND(P)$ są nazywane *podstawowymi kategoriami wiedzy P* . Jeżeli $Q \in R$ (tzn. Q jest relacją równoważności w U), to jej klasy równoważności nazywamy *pojęciami Q -elementarnymi wiedzy R* . Niech teraz $X \subseteq U$, a X niech będzie pewną relacją równoważności. Wtedy X jest zwane *R -definiowalnym* (lub *R -dokładnym*) jeśli jest sumą pewnych kategorii *R -podstawowych*. W przeciwnym przypadku, jest *R -przybliżone*.

Zbiór przybliżony określa się teraz w sposób przybliżony przypisując każdemu podzbirowi X w U oraz relacji równoważności F w R następujące dwa zbiory:

$$(a) \quad F_L X = \bigcup \{Y \in U \mid Y \subseteq X\}$$

$$(b) \quad F_U X = \bigcup \{Y \in U \mid Y \cap X = \emptyset\}$$

z których pierwszy jest *dolną aproksymacją*, a drugi *górną aproksymacją* zbioru X .

Dwa następujące pojęcia są teraz dla nas istotne: *jądro* (ang. core) i *konstrukt* (ang. construct). Są one związane z redukcją wiedzy, tzw. z wyrażaniem jej w najbardziej efektywny sposób. Niech więc R będzie znów rodziną relacji równoważności, a F niech będzie jednym z jej elementów. Wiedzy F jest *pomijalne* w R , jeśli

$$IND(R) = IND(R \setminus \{F\})$$

a w przeciwnym przypadku, F jest *niezbędne*.

Jeżeli teraz $Q \subseteq P$, gdzie P jest rodziną relacji równoważności, a ponadto $IND(Q) = IND(P)$, to Q jest reduktom P . Jądro jest z kolei zbiorem wszystkich niezbędnych relacji w P . Można pokazać, że jądro jest przecięciem wszystkich reduktów.

Mówiąc obrazowo, jądro składa się z takich klasyfikacji, które są najbardziej istotne z punktu widzenia dysponowanej wiedzy, czyli żadnej relacji w jądrze nie można usunąć bez zniekształcenia tej wiedzy. Z drugiej zaś strony, redukt podaje zbiór relacji, który jest wystarczający do scharakteryzowania danej wiedzy bez straty czegokolwiek istotnego.

3. ZASTOSOWANIE ZBIORÓW PRZYBLIŻONYCH DO PROGNOZOWANIA WYNIKÓW WYBORÓW

Podamy teraz ideę wykorzystania pewnych elementów teorii zbiorów przybliżonych do przewidywania wyników wyborów. Metoda jest ściśle związana z zastosowaniem podobnej metodologii w zbliżonej dziedzinie, a mianowicie z pracą Fedrizziego, Kacprzyka i Nurmiego (1993), w której zaproponowano zastosowanie teorii zbiorów przybliżonych do analizy procedur głosowań, głównie w sensie wyznaczenia zbioru kryteriów, jakie rzeczywiście różnicują te (często zbliżone do siebie) procedury. Najogólniej biorąc, w powyższej pracy wyznacza się redukt, czyli minimalny zbiór niezbędnych kryteriów, które dokonują takiego podziału, jak zbiór kryteriów w całości. Okazuje się zresztą, że w rozpatrywanym tam przypadku ten redukt jest wyznaczony jednoznacznie i jest równoważny jądru. A więc, procedury głosowań można zróżnicować stosując przynajmniej powyższy zbiór kryteriów.

Powyższą ideę można zastosować w analogiczny sposób do próby dokonania prognozy wyników wyborów. Kolejne kroki proponowanej w niniejszej pracy idei takiej procedury są przy tym następujące:

1. Wyznacza się zestaw kryteriów, które służyć będą do różnicowania poszczególnych partii, przy czym mogą tu występować kryteria bardzo różnych typów, jak np.:

- cechy programów mogące mieć znaczenie dla wyborców (np. świadczenia socjalne, wysokość podatków, minimalne pensje, renty i emerytury, preferencje w rozwoju branż przemysłowych itp.),
- cechy partii jako całości i jej przywódców, które mogą mieć wpływ na "image", jak np. umiejętność zjednania sobie środków masowego przekazu,
- sprawność struktur partyjnych,
- cechy kampanii wyborczej (np. aktywność i profesjonalność),

Jak więc widzimy wśród tych cech mogą być zarówno ilościowe jak jakościowe. Oczywiście, wybór tych cech jest tu raczej "sztuką" i zwykle do tego celu wykorzystuje się opinie ekspertów.

Każdą partię opisujemy teraz przez powyższy zestaw kryteriów, tworząc tzw. profil kryteriów partii.

2. Następnie wykorzystuje się opisaną metodę zbiorów przybliżonych do wyznaczenia

minimalnego zbioru kryteriów różnicujących poszczególne partie (reduktu).

3. Mając wyznaczony minimalny zbiór kryteriów różnicujących, należy teraz uzyskać informacje nt. istotności tych kryteriów z punktu widzenia głosowania na poszczególne partie. Zwykle trzeba tu zasięgnąć porady u ekspertów. Mogą oni przy tym podać albo jedynie informację jakościową, jak np. preferencje lub uporządkowanie kryteriów od najbardziej istotnego do najmniej istotnego, albo informację ilościową, jak np. intensywność takich preferencji.

4. Zakładając na początek informację jakościową, dokonuje się porównania kryteriów opisujących poszczególne partie z wyznaczonymi powyżej uporządkowaniami kryteriów i wyznaczamy uporządkowanie wśród partii od spełniających najwięcej z tych najbardziej preferowanych kryteriów do spełniających najmniej z nich. Powyższy wynik ma charakter jakościowy w tym sensie, że wyznacza partię (lub partie), które mają największe szanse na zwycięstwo w wyborach.

Wynik ilościowy, prognozę liczby uzyskanych mandatów, można natomiast uzyskać uwzględniając otrzymaną od ekspertów intensywność preferencji w zbiorze kryteriów i pojęcie odległości między profilami kryteriów dla poszczególnych partii (por. Fedrizzi, Kacprzyk i Nurmi, 1993). Jest to jednak dużo bardziej skomplikowane i będzie omówione w dalszych pracach.

4. UWAGI KOŃCOWE

W pracy przedstawiono ideę podejścia do prognozowania wyników wyborów z wykorzystaniem dodatkowych informacji i opinii ekspertów. Zastosowanie elementów teorii zbiorów przybliżonych pozwala na "inteligentną" analizę danych i uzyskanie w prosty sposób jakościowej prognozy. Nietrudno zauważyć, że istotnym elementem zaproponowanego podejścia są jakościowe i subiektywne oceny ekspertów, co jest jednak zgodne ze specyfiką rozpatrywanego zagadnienia.

LITERATURA

- Fedrizzi M., J. Kacprzyk i H. Nurmi (1993) How different are social choice functions: a rough set approach. W P. Eklund and J. Mattiła (Eds.): FIF'93 - Fuzziness in Finland, Abo Akademi, Ser. B, No 15, str. 25 - 38.
- Gambarelli G. (1994) Power indices for political decision making. *Annals of Operations Research* 51, str. 163 - 173.
- Hołubiec J. (1994) O pewnych zastosowaniach indeksów siły. W: *Metody i środki wspomaganie procesów decyzyjnych, BOS'93, PTBOiS*, Warszawa, str. 13-16.
- Hołubiec J. i J. Mercik (1992) *Techniki i tajniki głosowania*. Omnitech Press, Warszawa.
- Nurmi H. (1987) *Comparing voting systems*. Reidel, Dordrecht.
- Pawlak Z. (1982) Rough sets. *Int. J. of Inf. and Computer Sci.* 11, 341 - 356.
- Pawlak Z. (1991) *Rough sets: theoretical aspects of reasoning about data*. Kluwer, Dordrecht.

ISBN 83-85847-85-5

**W celu uzyskania bliższych informacji i zakupu dodatkowych egzemplarzy
prosimy o kontakt
z Instytutem Badań Systemowych PAN
ul. Newelska 6, 01-447 Warszawa
tel. 36-19-01 w. 241 e-mail: kotuszew@ibspan.waw.pl**