

Studiul voturilor senatorilor cu ajutorul algoritmului DVS

Militaru David-Mihai, 324AC

11 aprilie 2023

1 Introducere

Scopul lucrării este de a studia modalitățile de votare a senatorilor americani în cadrul celui de-al 116-lea Congres, printr-o abordare cât mai obiectivă. O astfel de abordare o reprezintă aplicarea tehnicilor de *data-mining*, precum DVS (Descompunerea valorilor singulare) asupra matricii alcătuite din voturile senatorilor. Astfel de modalități au fost folosite cu succes de-a lungul anilor de către oamenii de știință din domeniul politicii. Cu ajutorul algoritmului DVS, putem determina atât afilierea de partid a senatorilor, cât și caracteristicile proiectelor de lege asupra cărora s-a votat.

2 Descompunerea valorilor singulare

Teorema 1 *Orice matrice $A \in \mathbb{R}^{m \times n}$ poate fi factorizată astfel:*

$$A = USV^T, \quad (1)$$

unde $U \in \mathbb{R}^{m \times m}$ și $V \in \mathbb{R}^{n \times n}$ sunt matrici ortogonale, iar $S \in \mathbb{R}^{m \times n}$ este o matrice diagonală cu $r = \text{rang}(A)$ intrări pozitive.

Intrările matricii S , notate cu σ_i , pentru $i = 1, \dots, p$, cu $p = \min(m, n)$, se numesc **valorile singulare** ale matricii A și sunt ordonate descrescător ($\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$; $\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_p = 0$). Mai mult, coloanele $u_i \in \mathbb{R}^m$ ale matricii U se numesc vectori singulari stânga ai matricii A , în timp ce coloanele $v_i \in \mathbb{R}^n$ se numesc vectori singulari dreapta. Astfel, matricea A mai poate fi scrisă astfel:

$$A = \sum_{i=1}^r u_i \sigma_i v_i^T. \quad (2)$$

Teorema 2 *Valorile singulare nenule $\sigma_i, i = 1 : r$ ale matricii $A \in \mathbb{R}^{m \times n}$ sunt rădăcinile pătrate pozitive ale valorilor proprii nenule ale matricii simetrice pozitiv semidefinite*

$$B = AA^T,$$

numită și matrice de covarianță. Prin urmare, dacă $\lambda_1 \geq \dots \geq \lambda_r > 0$ sunt valorile proprii nenule ale lui B , atunci

$$\sigma_i = \sqrt{\lambda_i}.$$

Mai mult, coloanele matricei U sunt, de fapt, vectorii proprii ai matricei de covarianță AA^T , în timp ce liniile matricei V^T sunt vectorii proprii ai matricei $A^T A$.

3 Formularea matematică a aplicației

Folosind baza de date aferentă rezultatelor voturilor senatorilor, stocată într-un fișier de tip CSV, se construiește matricea $A \in \mathbb{R}^{m \times n}$, formată din câte o linie pentru fiecare dintre cei m senatori și câte o coloană pentru fiecare dintre cele n proiecte de lege asupra cărora s-a votat (în cazul de față, 101 senatori și 460 proiecte de lege). Elementul A_{ij} va avea valoarea $+1$ dacă senatorul i a avut un vot afirmativ la legea j , sau -1 dacă nu. În cazul în care respectivul senator a lipsit de la vot sau a ales să nu voteze, valoarea elementului va fi 0 .

Algoritmul DVS ne ajută să identificăm grupuri de senatori care au votat la fel asupra mai multor proiecte de lege. Astfel, matricea U rezultată este formată din vectorii proprii ai matricei de covarianță, deci coloanele acesteia reprezintă direcțiile de variație a tuturor valorilor din matricea de voturi. Mai mult, procentul de variație este dat de valoarea proprie corespunzătoare fiecărui vector în parte.

Utilizându-se relația dintre valorile singulare ale matricii de voturi și valorile proprii ale matricii de covarianță, identificăm, astfel, următoarele procentaje:

- u_1 : 52.31%
- u_2 : 28.02%
- u_3 : 2.32%
- u_4 : 1.53%
- u_5 : 1.13%

Asadar, aproximarea de rang 2 a matricei initiale, data de

$$A^{(2)} = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T$$

este o reconstituire fidelă a matricii initiale A , deci cele două coordonate cu care putem oferi predicții asupra modurilor de votare a senatorilor sunt date de primii 2 vectori, adică primele două componente principale.

4 Rezultate

4.1 Comportamentul senatorilor

Relatia de ortogonalitate dintre cei doi vectori principali ne permite sa plotam comportamentul senatorilor pe cele doua coordonate. Una dintre cele doua directii este numita "Partisan Coordinate" si reprezinta masura in care un senator a votat asemanator cu ceilalti membri ai partidului din care face parte. Cealalta directie, "Bipartisan Coordinate", reprezinta masura in care un senator a votat la fel cu majoritatea. Se obtine urmatorul grafic:

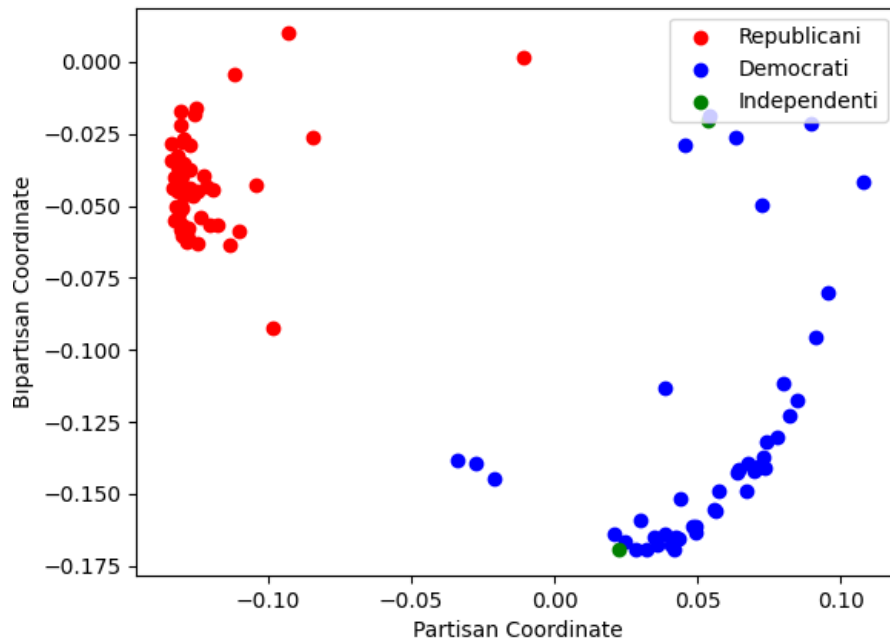


Figura 1: Modul de votare al senatorilor

Se observa ca:

- Democratii sunt complet separati de republicani, cu cateva mici exceptii
- Republicanii sunt foarte apropiati pe coordonata partizana si au valoare mult mai mare pe axa bipartizana decat democratii, de unde rezulta superioritatea lor numerica in senat (voteaza mai des la fel si, implicit, la fel cu majoritatea).

4.2 Reconstituirea fiecarui vot

Mai departe, vom construi aproximarea de rang 2 a matricii initiale si vom compara elementele matricelor intre ele. In matricea reconstituita, voturile fiecarui senator vor fi date de semnul elementelor respective. In figura 3, aratam procentul de voturi corect reconstituite pentru fiecare senator, punandu-se in evidenta **predictibilitatea** acestora.

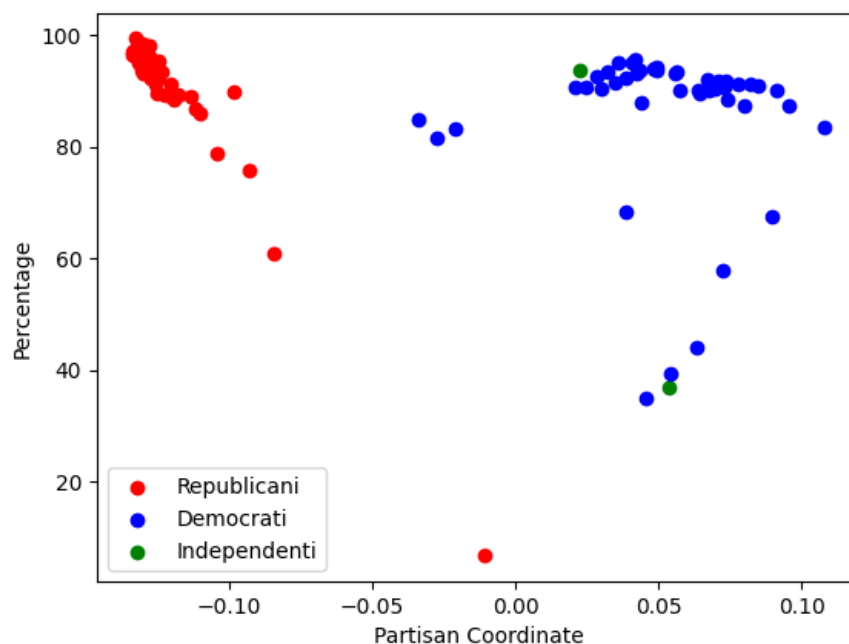


Figura 2: Predictibilitatea senatorilor

Se observa ca:

- Republicanii sunt mai predictibili decat democratii, datorita superioritatii lor si a extremismului fiecaruia (cel mai mare procentaj este de 99%)
- Un republican are predictibilitate foarte mica deoarece el a inlocuit un alt republican dupa un numar mare de proiecte de lege deja aprobate sau dezaprobrate, deci in matricea initiala, pe linia aferenta acestuia sunt multe elemente de 0 care si-au modificat valoarea odata cu aproximarea.

Se poate calcula, totodata, si numarul rezultatelor corect reconstituite. Adunandu-se elementele de pe fiecare coloana a matricei de rang 2 si comparandu-se cu elementele aferente ale matricei initiale, se constata ca au fost prezise corect rezultatele a 459 dintre cele 460 proiecte de lege.

4.3 Caracteristica voturilor

Putem calcula pozitia voturilor de-a lungul aceleiaşi direcţii, dar vom folosi primele doua linii ale matricei V^T .

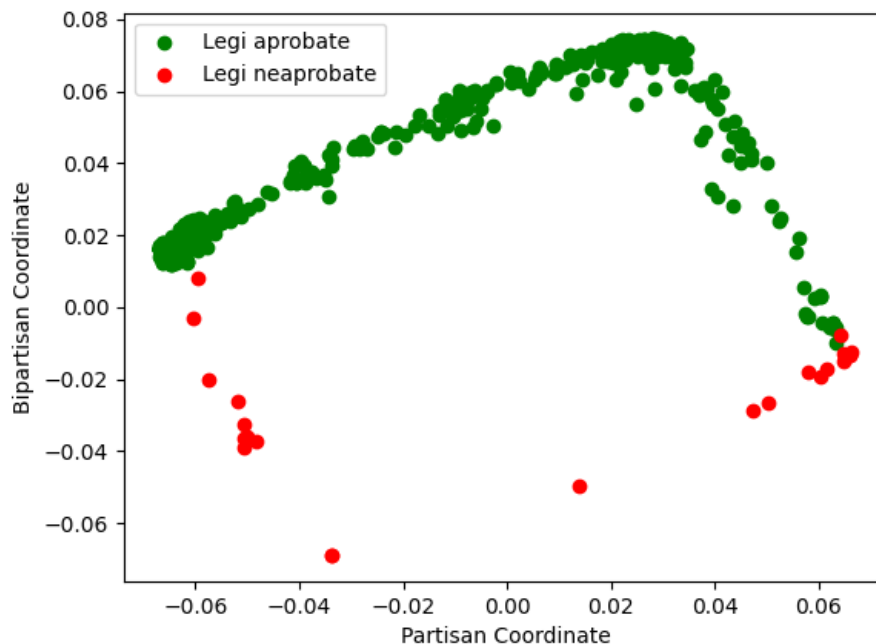


Figura 3: Predictibilitatea senatorilor

Cu cat procentul senatorilor care au votat o anumita lege scade, cu atat pozitia votului pe grafic se deplaseaza inspre stanga sau dreapta, in functie de caracteristica votului respectiv (interes republican sau democrat). Se observa ca:

- voturile cu coordonata bipartizana cea mai mare au fost aprobate, intrucat majoritatea a votat afirmativ;
- voturile cu coordonata partizana avand semnul coordonatei partizane a republicanilor au fost, in cea mai mare parte, aprobate, gratie superioritatii partidului republican;
- unele voturile cu coordonata partizana avand semnul coordonatei partizane a democratilor au fost dezaprobat, din cauza inferioritatii partidului democrat;
- voturile cu coordonata bipartizana negativa au fost dezaprobat.

Bibliografie

- [1] Martin, Porter, "The Extraordinary SVD"