

Comparação quantitativa da atuação parlamentar de partidos políticos utilizando a Análise de Componentes Principais

Resumo

Neste artigo propomos um método de comparação quantitativa entre as atuações de partidos políticos no poder legislativo. Este método utiliza como entrada dados de votações dos parlamentares em plenário, processa estes dados com o algoritmo Análise de Componentes Principais (PCA), e apresenta como resultado um gráfico bidimensional em que pontos representam partidos, e no qual a distância entre esses pontos reflete a semelhança entre os votos realizados pelos parlamentares.

1 Introdução

O Brasil é uma jovem democracia com um sistema político-partidário bastante fragmentado. Atualmente existem no Brasil X partidos políticos, sendo que pelo menos Y deles podem ser considerados pequenos partidos, pois não possuem qualquer representatividade no congresso nacional. Este quadro torna muito difícil para que o eleitor compreenda as diferenças práticas e ideológicas entre os partidos políticos. Desta forma, tornam-se valiosos os procedimentos que permitam estabelecer comparações entre os partidos, em particular as comparações objetivas, por não dependerem de julgamento subjetivo dos analistas.

Dentre os agentes políticos de interesse que atuem em nome de partidos políticos, os parlamentares apresentam uma particularidade que os tornam mais fáceis de serem analisados com uma abordagem quantitativa. Dois ou mais parlamentares podem ser objetivamente comparados analisando-se suas preferências em votações de proposições legislativas. Dois parlamentares são “semelhantes”, ou “próximos”, se tendem a votar igual na maioria das vezes.

Dois executivos diferentes não podem ser comparados da mesma forma, pois dois executivos diferentes não resolvem exatamente o mesmo problema nas mesmas condições. Discursos ideológicos e programáticos podem ser comparados, mas dificilmente de forma quantitativa.

A comparação entre dois parlamentares pode ser feita de forma bastante simples, como por exemplo medindo a porcentagem de acordo entre eles sobre um conjunto de votações. No entanto, comparar três ou mais parlamentares concomitantemente e dispô-los em algum tipo de escala não é trivial. Da mesma forma, modelar comparações entre partidos com base nos valores das votações também requer algum procedimento mais sofisticado.

Neste estudo utilizaremos a Análise de Componentes Principais (PCA) como ferramenta matemática para realizarmos a comparação de um dado conjunto de votações, a fim de estabelecermos semelhanças quantitativas entre os partidos políticos que atuaram neste conjunto de votações.

Nossas principais contribuições são:

- uma modelagem matemática de votações parlamentares que fornece a entrada necessária ao algoritmo PCA;
- evidências de que é possível utilizar o PCA para comparar quantitativamente a atuação legislativa de partidos políticos.

Este artigo se organiza da seguinte forma: na Seção 2 apresentaremos outros métodos quantitativos utilizados na Ciências Políticas para comparação de votações parlamentares; na Seção 3 forneceremos uma breve descrição sobre o funcionamento da PCA em termos de entrada requerida e saída gerada, enquanto que na Seção 4 descreveremos nossa modelagem matemática para gerar um conjunto de vetores a partir de dados de votações, de forma que esses vetores sejam entrada da PCA; na Seção 5 descreveremos os resultados obtidos com a aplicação da técnica proposta em dados reais de votações parlamentares; por fim, na Seção 6 apresentamos nossas conclusões.

2 Trabalhos relacionados

Nominate

Problemas do nominate

3 A Análise de Componentes Principais

A Análise de Componentes Principais (PCA) é o método estatístico mais popular para redução dimensional de grandes conjuntos de dados [1]. A PCA

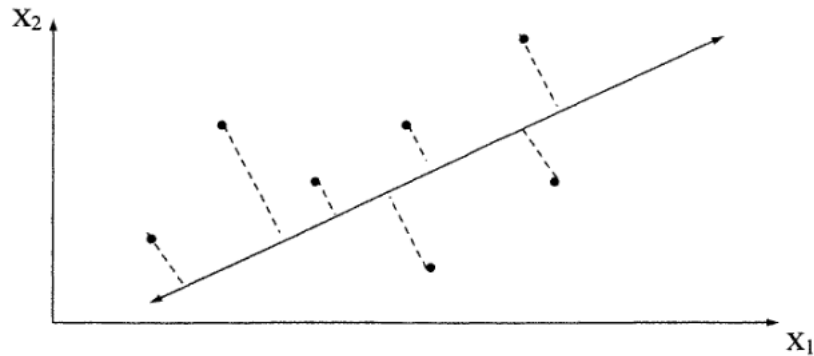


Figura 1: Exemplo de PCA aplicado na redução de 2 dimensões para 1 dimensão. Retirado de “Data Mining, Concepts, Models, Methods, and Algorithms” [1]

realiza uma *transformação linear* que recebe como entrada um conjunto de N vetores¹ de dimensão M e produz um outro conjunto de N vetores de dimensão M , de tal forma que a maior parte da informação nos vetores de saída fica acumulada nas primeiras dimensões. *Informação* aqui se refere às diferenças entre os vetores de um dado conjunto. Mais precisamente, a primeira dimensão da saída da PCA possuirá a maior variância entre seus elementos, enquanto que a segunda dimensão possuirá a segunda maior variância, e assim por diante. Esta propriedade nos permite reduzir a dimensão de um conjunto de dados com o mínimo de perda de informação possível.

Um exemplo didático é apresentado no livro “Data Mining, Concepts, Models, Methods, and Algorithms” [1], em que um conjunto de dados com duas dimensões (pontos em um gráfico bidimensional) é reduzido para um conjunto de dados com apenas uma dimensão (pontos em uma reta). Na figura podemos observar como o eixo que determina a primeira dimensão do resultado da PCA garante a máxima variância entre os dados. Qualquer outro eixo escolhido deixaria as projeções dos pontos do gráfico sobre o eixo mais próximas entre si do que podemos ver no eixo adotado na figura.

Mostrar outros contextos em que o PCA é usado

¹Um vetor é um conjunto ordenado de valores, em que cada posição é chamada de dimensão. Um exemplo de um vetor de 3 dimensões é $v = [10, 15, 12]$.

4 Método proposto

Conforme explicado na Seção 3, a utilização da PCA requer como entrada um conjunto de vetores. Para que possamos utilizar a PCA para comparar partidos políticos objetivamente com base nos votos dados por parlamentares, precisamos então de uma modelagem matemática que gere um conjunto de vetores representativos de um dado conjunto de votações.

Nesta seção apresentaremos nossa proposta para a criação de *vetores de votações* que sirvam de entrada para a PCA. Um vetor de votação modela as preferências de um partido político em um conjunto de votações. Dessa forma, aplicaremos a PCA com o objetivo de reduzir a dimensionalidade dos vetores de votação, tendo como saída de nosso algoritmo um conjunto de vetores de duas dimensões que podem ser representados como pontos em um gráfico bidimensional. Neste gráfico bidimensional gerado, cada ponto representará um partido, e quanto mais próximos um grupo de partidos estiverem no gráfico, mais semelhantemente terão esses partidos votado nas votações analisadas.

No caso mais simples, em que o partido possui apenas um parlamentar em uma dada casa legislativa, a i -ésima posição do vetor de votações é preenchido com um valor numérico correspondente ao valor do voto desse parlamentar na i -ésima votação do conjunto analisado. Esse valor numérico, correspondente ao voto de um parlamentar d em uma votação v , é definido pela *função opção*, $op(d, v)$, que é definida como se segue:

$$op(d, v) = \begin{cases} 1 & , \text{ se parlamentar votou } \textit{sim} \\ -1 & , \text{ se parlamentar votou } \textit{não} \\ 0 & , \text{ em algum outro caso} \end{cases}$$

Os “outros casos” além do sim e do não podem consistir na abstenção, ausência, ou mesmo obstrução. Todos esses casos representam uma impossibilidade em verificar a opinião do votante sobre a votação.

Leo ► *TODO: enumerar suposições/premissas* ◄

Formularemos agora a definição de um vetor de votações pertencente a um partido. Seja:

- C a casa legislativa sob análise;
- p um partido político com representatividade em C ;
- D_p o conjunto dos parlamentares do partido p em C ;
- $V = \{v_1, v_2, \dots, v_n\}$ um conjunto de votações ocorridas em C .

Então o vetor de votações do partido p , representado por vv_p , será dado por:

$$vv_p[i] = \frac{\sum_{d \in D_p} op(d, v_i)}{|D_p|}$$

A fórmula acima expressa que a i -ésima posição do vetor de votações do partido p possui como valor a média aritmética dos valores da função opção para todos os parlamentares do partido p na votação v_i .

Exemplificando, assumamos o seguinte cenário: todos os parlamentares de p votaram *sim* na votação v_1 ; todos os parlamentares de p votaram *não* na votação v_2 ; todos os parlamentares de p faltaram à votação v_3 ; e por fim, metade dos parlamentares de p votaram *sim* na votação v_4 , mas a outra metade votou *não* nessa mesma votação v_4 . Nesta situação o vetor de votações do partido p pode ser representado por $vv_p = [1, -1, 0, 0]$.

O exemplo anterior evidencia como a abstenção e a ausência apresentam as mesmas consequências em nossa modelagem. Essa mesma consequência pode ser semanticamente interpretada como uma “falta de opinião formada do partido sobre o assunto”. Claro que na realidade os parlamentares podem efetuar manobras, como baixar o quórum para impedir a votação, situação na qual se torna difícil entender se o partido seria a favor do *sim* ou do *não*. Desta forma, assumimos essa simplificação para permitir que nosso método se mantenha puramente quantitativo, sem que seja necessário a inserção do julgamento de analistas.

Os vetores de votação gerados pelo procedimento explicado serão fornecidos como entrada da Análise de Componentes Principais. Ao aplicar a PCA sobre esses vetores de votações teremos como saída um novo conjunto de vetores de votação, mas com suas diferenças acumuladas nas primeiras dimensões. Dessa forma, o resultado final é apresentado em um gráfico de duas dimensões, cujos pontos possuem coordenadas correspondentes às duas primeiras dimensões dos vetores gerados pela PCA.

Na próxima seção descreveremos os resultados obtidos aplicando-se a PCA com a modelagem aqui apresentada a dados reais de votações parlamentares.

5 Resultados

Mostrar gráficos e gerados

6 Conclusões

Referências

- [1] Mehmed Kantardzic. Section 3.4 Principal Componente Analysis. In *Data Mining, Concepts, Models, Methods, and Algorithms*. John Wiley & Sons, INC., 2003.