Análise de Agrupamento das Cartas do Tipo Criatura em *Magic: The Gathering*

Felipe Neres Silva Bezerra e Patrícia

Universidade Cruzeiro do Sul

RESUMO

Este texto descreve a análise de agrupamento aplicada a um componente do jogo de cartas *Magic: the Gathering*, especificamente um tipo de cartas denominada como *criaturas*. Os grupos obtidos foram baseados nos padrões de variáveis encontradas em todas as observações: *força*, *resistência* e *custo*. Cinco métodos de agrupamento hierárquico foram aplicados e a relação entre dispersão interna dos grupos e número de grupos mantidos foi avaliada; o método de Ward mostrou-se o mais satisfatório neste quesito e o estudo prosseguiu com cinco grupos.

Segundo a predominância dos valores das variáveis, os grupos foram nomeados como criaturas **frágeis**, **defensivas**, **intermediárias**, **ofensivas** e **descomunais**. Além de servir de ponto de partida para diversos outros testes, os grupos formados respeitaram e reforçaram preceitos estratégicos do jogo, considerando suas suposições implícitas.

**Palavras-chave**: Análise de agrupamento, *Magic: the Gathering*.

ABSTRACT

This text describes the clustering analysis applied to a component of the *Magic: the Gathering* card game, specifically a kind of cards called *creatures*. The clusters obtained were based on the patterns of variables found in all the observations: *strength*, *resistance* and *cost*. Five hierarchical clustering methods were applied and the relationship between dispersion within the clusters and number of clusters kept was evaluated; Ward's method proved to be the most satisfactory in the study proceeded with five clusters.

According to the predominance of variable values, the clusters were named as **fragile**, **defensive**, **intermediate**, **offensive**, and **oversized** creatures. Besides serving as a starting point for several other tests, the formed clusters regarded and reinforced strategic precepts of the game, considering their implicit assumptions.

**Keywords**: Clustering analysis, *Magic: the Gathering*.

1. INTRODUÇÃO

*Magic: the Gathering* (também conhecido apenas como *Magic*) é um jogo de cartas colecionável no qual cada jogador – normalmente dois, mas variando de acordo com o modo de jogo – utiliza seu próprio baralho em turnos alternados com o objetivos principais, entre outros, de reduzir a pontuação denominada como “*pontos de vida*” do adversário, inicialmente somando um total de 20 pontos, a zero, ou ainda conduzir a partida ao momento em que o adversário deve sacar uma carta do próprio baralho mas este já estiver esgotado[[1]](#footnote-1). Como o jogador é o único responsável pelas cartas que possui à sua disposição ao longo da partida, as habilidades dos jogadores postas em teste não se resumem à utilização correta das cartas no momento mais oportuno do início ao fim de cada partida, mas se estende desde a seleção das cartas que comporão seu baralho pessoal.

O número de cartas diferentes à disposição do jogador para compor seu baralho varia de acordo com o formato[[2]](#footnote-2) estabelecido para a partida, mas os tipos de cartas disponíveis são sempre os mesmos: *artefatos*, *criaturas*, *encantamentos*, *feitiços*, *mágicas* *instantâneas*, *planinaltas* e *terrenos*. Como cada carta carrega em si parte das regras, modificando as regras básicas ou acrescentando regras novas, há informações impressas nelas que são comuns para que se mantenha a coerência ao longo do jogo. Isso mantém a estrutura das informações nas cartas de forma similar, mas não completamente idêntica, assegurando padrões apenas entre cada tipo de carta. Como destaque dentre as divergências nas estruturas das cartas, as *criaturas* e os *planinaltas* possuem informações que não constam nos demais tipos: as *criaturas* possuem valores de *força* e *resistência*, enquanto que *planinaltas* possuem valores iniciais de *lealdade*. Ambos os tipos, por possuírem mais dados quantificáveis, são mais assertivos. Porém, o número de cartas de planinaltas diferentes não é tão expressivo quanto o de criaturas.

Tendo em vista estas considerações preliminares, com intuito de aplicar o método estatístico de análise de agrupamento (*cluster*) a um assunto familiar aos autores, o objetivo do presente relatório será organizar em grupos as diversas cartas do tipo *criatura* do jogo *Magic: the Gathering*. A sessão a seguir tratará da seleção de variáveis, determinação de medidas de dissimilaridade e esclarecimentos acerca dos métodos de agrupamento. Em seguida, os resultados serão abordados. Por fim, serão feitas considerações finais e impressões sobre o processo e os resultados.

1. METODOLOGIA

Dentre os vários momentos em uma parida de Magic, um dos mais comuns é o confronto entre criaturas. Embora isso possa ocorrer entre um número indefinido de criaturas, utilizar o confronto entre duas criaturas é a forma mais simples de exemplificar essa ação demonstrar a relevância dos atributos escolhidos para o agrupamento.

Em determinado momento na vez de um jogador, ele pode declarar um ataque com quantas de suas criaturas em jogo quiser. O oponente, por sua vez, pode escolhe qualquer número de suas criaturas para bloquear o ataque. Quando uma criatura ataca e outra defende, os valores de força e resistência são comparados de forma cruzada: dentre ambas, a que possuir resistência igual ou inferior à força da criatura oposta é derrotada. Assim, as criaturas com maior força tem mais chances de derrotar os as criaturas adversárias, enquanto que as que possuem maior resistência tem mais chance de não serem derrotadas em confrontos.

À parte das características das criaturas, outro valor recorrente na grande maioria das cartas de Magic é o custo. Cartas que fornecem maiores vantagem costumam ter maiores exigências para entrar em jogo, o que se traduz em maior custo, que é pago com *Mana*, que se trata do recurso que os jogadores precisam gerenciar a cada turno. Embora o custo seja em parte explicado pela força e pela resistência da criatura, há ainda vantagens advindas de efeitos impressos na carta que alteram ou adicionam regras à partida e que podem torna-la mais vantajosa, implicando em custos mais elevados para que se utilize criaturas com atributos não tão relevantes. *A priori*, todo o custo que não pode ser explicado pela força e pela resistência da criatura deveria ser explicada pelo seu efeito impresso em texto.

Com um total de 8370 observações disponíveis na base de dados[[3]](#footnote-3), medidas em três variáveis discretas que variam entre 0 e 16, diversas criaturas, embora não igual para o jogo, foram tratadas como reincidências quando possuíam a mesma exata combinação de valores que quaisquer outras criaturas. Para resolver a questão de reincidência, foi realizado um tratamento para que houvesse somente um elemento com cada combinação de valores, considerando reincidências como a mesma observação (criaturas iguais em valores serão pertencerão ao mesmo grupo). Após o tratamento, ainda restaram 380 observações, o suficiente para que se prosseguisse com as análises sem demandar muito do poder computacional disponível. Como as observações se distribuem de forma discreta até um máximo próximo em todas as variáveis, qualquer processo de padronização ou escalonamento dos dados foi dispensável. Ademais, criaturas que não apresentavam valores fixos (estabelecidos apenas ao longo da partida) também não foram consideradas.

{Gráfico tridimensional, ou três gráficos bidimensionais}

Figura 1: disposição das observações ao longo das variáveis mensuradas.

As três variáveis supracitadas são quantitativas discretas, possibilitando utilizar tanto a métrica de similaridade (correlação) quanto de dissimilaridade (distância). Como a relevância dos grupos não se dará pelo comportamento das variáveis em cada observação, foi escolhida a métrica de distâncias; especificamente a distância euclidiana, por possibilitar uma demonstração gráfica mais intuitiva em comparação às demais. A distância de Mahalanobis seria preferível quando observa-se multicolinearidade (HAIR et al, 2009, p. 442 e 447), o que não é o caso, conforme ilustrado na figura a seguir. Também não ocorre problema de representatividade, pois foi utilizada toda a polução no estudo.

{Gráfico de correlações}

Figura 2: correlação entre as variáveis.

Os procedimentos de agrupamento utilizados restringiu-se aos métodos hierárquicos, devido à simplicidade e à afinidade com medidas de similaridade. Todos os algoritmos de agrupamento foram computados com o auxílio do software R versão \*, com os pacotes “*tidyverse*”, “*cluster*”, “*NbClust*”, “*pvclust*”, “*ggplot2*” e “*plotly*”.

1. RESULTADOS

O processo de agrupamento, grosso modo, consiste em classificar observações com valores de variáveis mais próximos sem qualquer hipótese a priori. A proximidade das observações é definida pela métrica de similaridade (distância ou correlação), enquanto o processo de junção e classificação fica a cargo do método de agrupamento. Todos os métodos avaliados são processos hierárquicos; ou seja, cada grupo pode ser considerado parte de um grupo maior, sendo possível identificar as ramificações desde cada observação sendo considerada seu próprio grupo até um único grupo que abrange todos os dados; em determinado ponto da ramificação é estipulado o número ideal de grupos para que se prossiga com a análise. Os métodos avaliados foram o método de médias ponderadas (*Weighted Pair Group Method with Averaging* – WPGMA), médias não ponderadas (*Unweighted Pair Group Method with Averaging* – UPGMA), centroides ponderados (*Weighted Pair Group Method with Centroid* – WPGMC), centroides não ponderados (*Unweighted Pair Group Method with Centroid* – UPGMC) e método *Ward*.

Uma forma de medir o quão próximas as observações de um mesmo grupo se encontram é através da soma das dispersões dentro do grupo elevadas ao quadrado (*within sum of squares* – WSS). Em todos os métodos, quanto mais grupos são formados, menos observações cada grupo conterá e, com isso, menor será a dispersão total dentro de cada grupo; logo, o número máximo de grupos – quando cada observação é seu próprio grupo – garante o menor WSS, entretanto, isso equivaleria a não haver agrupamento. Tendo em vista o conflito de escolha entre um número gerenciável de grupos e manter observações de um mesmo grupo mais similares o possível, é exigido do pesquisador que encontre o número ideal de grupos. Para tal, foi utilizada análise gráfica da relação entre WSS e a quantidade K de grupos, denominado *elbow plot*. O ponto ideal é aquele no qual um grupo a menos elevaria em demasia o WSS, mas um grupo a mais levaria a uma redução irrelevante do WSS.

{Gráfico de cotovelo}

Figura 3: *elbow plot*.

Para todas as quantidades de grupos dispostas no gráfico-cotovelo, o método de Ward apresenta os menores valores de WSS – ou seja, organizou as observações em grupos mais concisos. Isso se explica pelo próprio processo do método: enquanto os métodos de média e centroide se baseiam na medida de similaridade sugerida, o método de Ward agrega cada par de grupos de modo a minimizar a soma dos quadrados de cada variável (HAIR et al, 2009, p. 452). Seguindo com uma avaliação visual do *elbow plot*, a interpretação do agrupamento segue com o total de cinco grupos.

{Número de observações em cada grupo}

Tabela 1: distribuição das observações nos grupos.

{Gráfico 3D com os 5 grupos}

Figura 4: disposição das observações ao longo das variáveis mensuradas, destacadas as pertenças de cada grupo.

Nota-se que o posicionamento dos grupos deu-se por blocos. O grupo 1 encontra-se mais próximo a origem dos eixos; o grupo 3 situado mais ao centro do gráfico, considerando os valores máximos das variáveis como as fronteiras do gráfico; os grupos 2 e 4 similares quanto ao eixo do custo, mas um situado em valores maiores no eixo da resistência e outro no eixo da força, respectivamente; e o grupo 5 composto pelas observações mais distantes considerando todos os eixos.

{Boxplots dos grupos}

Figura 5: *boxplots* dos grupos, separados por variáveis.

Se o custo é determinado em função da força, da resistência e de outras regras não quantificáveis – características exógenas ao modelo – promovidas por cada criatura, algumas suposições podem ser elaboradas a respeito do comportamento destes cinco grupos.

O grupo 1 é composto por criaturas com baixos atributos de força e resistência, mas com custos mais dispersos, o que sugere que as vantagens oriundas de efeitos destas criaturas impactam em seus custos de formas diversas. Ou seja, este é o grupo das criaturas mais **frágeis**, com ou sem efeitos relevantes.

Ao grupo 2 pertencem as criaturas cuja força é pouco relevante na determinação do custo, sendo estas mais **defensivas**.

No grupo 3 estão as criaturas **intermediárias**, tanto com relação ao custo quanto à força e à resistência.

O grupo 4 comporta-se como a contraparte do grupo 2: custo similar (ligeiramente superior), mas justificado majoritariamente pela força, o que as qualifica criaturas **ofensivas**.

No grupo 5, por fim, encontram-se as criaturas **enormes**, com valores de força e resistência superiores; em cada observação, força e resistência evoluem de forma linear neste grupo, como pode ser visto no gráfico tridimensional que dispõe o posicionamento geográfico dos grupos. Em relação ao custo, embora o grupo se destaque com os maiores níveis nesta variável, observa-se grande concentração em níveis inferiores, próximos ao que é encontrado no grupo 3; isso pode ser explicado pela existência de “*downsides*”: efeitos de alteração de regras da própria criatura que acrescente dificuldade à parte do custo, consequências negativas ou restrições quanto ao seu uso.

1. CONSIDERAÇÕES FINAIS

Há muitas informações não quantificáveis e diversas cartas possuem apenas esse tipo de informação além do custo, o que restringiu a pesquisa a uma parcela restrita de todo o universo do jogo. Ainda assim, a assertividade dos resultados pode ser considerada satisfatória, já que os grupos formados puderam ser explicados segundo a estrutura do jogo.

Como a análise de agrupamento é não inferencial, mas exploratória, serve como ponto de partida para outros testes estatísticos, como os sugeridos: pode-se aplicar testes de regressão linear para explicar o custo em função da força e da resistência em cada grupo; a aplicação de análise de variância (ANOVA) para verificar o quão significante é a diferença entre os grupos; ou ainda, acrescentando outras variáveis categóricas, como os subtipos das criaturas ou sua cor, e aplicando ao banco de dados completo (com todas as 8370 criaturas ao invés de 380 combinações de valores), é possível testar a distribuição ao longo das combinações de categorias com o teste chi².

Com a análise de agrupamento foi possível estabelecer classificações implícitas, que comumente são identificadas intuitivamente, tornando mais claro o panorama geral do Magic: the Gathering, principalmente para aqueles que ainda não estão familiarizados com todas as opções de estratégias disponíveis. Tal abordagem pode, ainda, servir como um dentre os diversos pontos da análise da estrutura de baralhos.

REFERÊNCIAS

HAIR JR., J. F., et al. **Análise Multivariada de Dados**. 6ª ed. Porto Alegre: Bookman. 2009.

GARFIELD, R.; TABAK, M. **Magic: the Gathering** – Manual de Regras Básicas. Renton: Wizards of the Coast. 2013.

1. Regras básicas de *Magic: the Gathering* disponíveis em: <http://media.wizards.com/images/magic/resources/rules/PT_MTGM14_Rulebook_Web.pdf>. Acesso em 25 de abril de 2019. [↑](#footnote-ref-1)
2. Diversos formatos de jogo disponíveis em: <https://magic.wizards.com/pt-br/content/formatos>. Acesso em 25 de abril de 2019. [↑](#footnote-ref-2)
3. Dados disponíveis em: <https://mtgjson.com/files/all-cards/>. Acesso em 18 de março de 2019. [↑](#footnote-ref-3)