

Uma algoritmo de agrupamento heterogêneo para formação de grupos de aprendizagem

Daniel F. dos Santos, Luziane F. de Mendonça, Marcello G. Teixeira

Depto de Ciências de Computação - DCC, UFRJ,

CP 68530, CEP 21941-590, Rio de Janeiro, RJ

E-mail: sfdanielrg@gmail.com, luziane@dcc.ufrj.br, marcellogt@dcc.ufrj.br

Resumo: *A maioria dos algoritmos desenvolvidos para a realização de agrupamento baseia-se nas semelhanças entre os elementos dos grupos. Entretanto, para a formação de grupos de estudos mais eficientes é preciso uma mudança no critério de escolha dos elementos. Pesquisas recentes revelam que grupos com indivíduos de características diferentes são mais eficientes ao promover o aprendizado de seus componentes.*

Neste trabalho, uma nova técnica de agrupamento é apresentada destinada à resolução do problema de formação de grupos como elementos heterogêneos. Os resultados mostram a eficiência do algoritmo proposto, no sentido de que os agrupamentos encontrados apresentaram alta heterogeneidade entre os elementos de cada grupo, e ao mesmo tempo a homogeneidade desejada entre os grupos. Além disso, a utilização em sala de aula dos agrupamentos gerados pelo algoritmo corroboraram com a hipótese das pesquisas, aumentando o rendimento dos alunos na aprendizagem colaborativa.

Palavras-chave: *Técnicas de Agrupamento, Grupos Heterogêneos, Aprendizado Colaborativo.*

1 Introdução

A necessidade de propor uma solução para o problema de criterizar a escolha dos componentes de grupos de estudos colaborativos, segundo um objetivo, nos remete ao estudo das técnicas de agrupamento.

Análise de agrupamento, ou *clustering*, é o nome dado para o grupo de técnicas computacionais cujo propósito consiste em separar objetos ou indivíduos em grupos, baseando-se nas características que estes possuem. Um grupo (*cluster*) é uma coleção de objetos de dados que são similares dentro de um mesmo grupo e são dissimilares entre *clusters* distintos [8].

O critério baseia-se normalmente em uma função (métrica) que afere a dissimilaridade, a qual recebe dois objetos e retorna a distância entre eles. Os grupos determinados por uma métrica de qualidade devem apresentar homogeneidade interna e heterogeneidade externa. Isto quer dizer que os elementos de um determinado conjunto devem ser mutuamente similares e, preferencialmente, muito diferentes dos elementos de outros conjuntos.

O estudo de agrupamento é uma ferramenta útil para a análise de dados em muitas situações diferentes. A análise de agrupamento tem aplicações em situações simples do dia-a-dia (desde o início da infância, aprende-se a distinguir entre gatos e cachorros, ou plantas e animais, melhorando continuamente esquemas de classificação subconscientes) assim como em diversas áreas, como Biologia, Pesquisa de Mercado, Processamento de Imagens, Reconhecimento de Padrões, Geografia, e muitas outras [8].

Os agrupamentos podem ser realizados e classificadas de várias formas: agrupamento hierárquico/particionado (dependendo se há intersecção entre os grupos), exclusivo/sobreposto/fuzzy (dependendo do tipo pertinência de cada elemento ao seu grupo), completo/parcial (dependendo se todos os objetos são alocados a grupos ou não), etc. [17].

Infelizmente o problema de agrupamento apresenta uma complexidade de ordem exponencial. Isto quer dizer que métodos de força bruta, como enumerar todos os possíveis grupos e escolher a melhor configuração, não são viáveis. Por exemplo, se quisermos separar 100 elementos em 5 grupos, existem $5^{100} \approx 10^{70}$ possíveis agrupamentos. Mesmo em um computador que possa testar 10^9 configurações diferentes por segundo, seriam necessários mais de 10^{53} anos para terminar a tarefa (cerca de 40 ordens de grandeza maior do que a idade da Terra).

Há inúmeras técnicas de agrupamento destinadas a diferentes tipos de problemas. Um dos algoritmos de agrupamento de dados mais usado e conhecido na literatura é o K-Means, que foi apresentado por J.B. MacQueen [13] em 1967. Este algoritmo tenta fornecer uma classificação de acordo com os próprios dados, sendo esta feita por similaridade (minimiza a distância acumulada de todos os elementos ao centro dos seus grupos). Outros algoritmos conhecidos são: K-Medoides, PAM (*Partitioning Around Medoids*), CLARA (*Clustering LARge Applications*), CLARANS (*Clustering Large Applications based upon RANdomized Search*) [11], etc.; todos esses são derivações do K-Means e tem o mesmo objetivo (agrupamento por similaridade); outro método bastante utilizado é o K-Center, o qual é semelhante ao K-Means, entretanto este minimiza a máxima distância entre os elementos de um mesmo grupo ao seu centro. Todos os métodos citados acima agrupam os objetos ao ponto mais próximo, ou seja, agrupam por similaridade (elementos homogêneos).

Analisando os métodos de agrupamento presentes na literatura, pode-se observar que a grande maioria (se não a totalidade) é reservada aos problemas de agrupamentos de elementos semelhantes. O problema abordado neste trabalho difere das abordagens tradicionais de clusterização, uma vez que necessita formar grupos de elementos heterogêneos.

Muitos pesquisadores dedicaram seus trabalhos para provar que grupos homogêneos e grupos heterogêneos apresentam interação e respostas diferentes às atividades propostas. Parker (1995)[15] enfatiza que nos grupos mais heterogêneos as pessoas têm a oportunidade de aprender novos comportamentos, de desenvolver habilidades interpessoais e de soluções de problemas, de desenvolver a capacidade de se adaptar a diferentes contextos e também compreender os diversos valores, crenças e estilos de trabalho das pessoas com quem trabalha.

Com o objetivo de criar um sistema de recomendação de grupos de estudos mais eficientes, acredita-se que a composição desse grupo deva respeitar uma discrepância mínima entre seus elementos. Essa diferença garante maior eficiência no objetivo da aprendizagem colaborativa, bem como no resultado final do objeto estudado. Neste contexto, o objetivo principal é criar grupos priorizando a heterogeneidade de seus membros e a homogeneidade entre os grupos.

Para tanto, na Seção 2 o problema estudado é apresentado, e o algoritmo proposto para sua resolução é descrito na Seção 3. Os resultados numéricos obtidos são expostos na Seção 4. Por fim, na Seção 5, algumas conclusões são apresentadas, assim como possibilidades de continuação deste trabalho.

2 O Problema

Nas últimas décadas, muitos pesquisadores têm se dedicado ao estudo sobre a forma ideal de composição de grupos de estudo/trabalho. As características individuais, as diferenças existentes entre indivíduos pertencentes ao mesmo grupo e as influências dessas diferenças sobre as interações humanas, sobre o funcionamento e desempenho dos grupos, bem como sobre o contexto nos quais se inserem, é um fenômeno que vem sendo estudado por vários pesquisadores desde o início do século XX (Asch, 1977 [1]; Lewin, 1948 [12]; Triandis, Hall, Ewen, 1965 [18]; Shaw 1981 [16]; Watson, Kumar, Michaelsen, 1993 [19]; Jehn, Northcraft, Neale, 1999 [9]; Jung, Sosik, Baik, 2002 [10]; etc.).

Com o avanço do Ensino a Distância (EaD) tornou-se muito mais frequente a prática de reunir grupos de estudos para realizar atividades dos cursos em questão. Foram realizadas várias pesquisas e desenvolvidas novas metodologias de ensino que fornecem instrumentos para o auxílio da aprendizagem em grupos. Porém, uma questão ainda permanece: como formar

os grupos? Atualmente, escolas de formação orientam educadores a criarem grupos segundo idade, sexo, proximidade, afetividade, enfim, características que resultam em grupos homogêneos segundo esses critérios. Além desses critérios não terem relação com o objetivo da atividade que o educador está administrando, os grupos homogêneos proporcionam uma troca de experiências com pouca qualidade entre os alunos.

Melser (1999)[14] mediu a auto-estima de alunos superdotados em grupos homogêneos e os comparou com a auto-estima daqueles que trabalham em um grupo heterogêneo. A avaliação da auto-estima dos dois grupos difere significativamente quando comparados. Enquanto os alunos superdotados trabalhando em grupo heterogêneo tiveram um aumento da auto-estima, os alunos superdotados que trabalharam em um grupo homogêneo tiveram uma redução da mesma.

Dada a necessidade de gerar grupos de forma diferente do que a literatura oferece nos métodos conhecidos, este trabalho propõe um novo método de agrupamento baseado na resolução de um problema de otimização.

Considere X um possível agrupamento, ou seja, $X = \{G_1, \dots, G_g\}$ onde cada G_i é um grupo com t elementos.

Neste trabalho, a heterogeneidade de cada grupo G_i é mensurada de forma proporcional a menor discrepância entre dois elementos desse grupo. Portanto, a heterogeneidade mínima obtida por um conjunto de grupos pode ser aferida como

$$R(X) = \min_{i=1, \dots, g} \min_{u, v \in G_i} \|u - v\|.$$

Por outro lado, a homogeneidade entre os grupos pode ser mensurada de forma inversamente proporcional ao valor da maior discrepância entre dois pontos médios desses grupos. Portanto, a homogeneidade mínima obtida por um conjunto de grupos pode ser aferida de forma inversamente proporcional a

$$M(X) = \max_{i=1, \dots, g} \max_{j=1, \dots, g} \left\| \sum_{u \in G_i} \frac{u}{t} - \sum_{v \in G_j} \frac{v}{t} \right\|.$$

O problema estudado visa um agrupamento X de forma a contemplar dois objetivos: máxima heterogeneidade de cada grupo e máxima homogeneidade entre eles; podemos definir uma função f que combina de forma ponderada os dois objetivos

$$f(X) = \lambda R(X) - (1 - \lambda)M(X). \tag{1}$$

Logo, o objetivo do problema é encontrar um agrupamento X que resolva o problema

$$\max f(X). \tag{2}$$

3 Algoritmo Proposto

Conforme mencionado anteriormente o objetivo consiste na formação de grupos de mesmo tamanho t , homogêneos, com composição o mais heterogênea possível.

O algoritmo proposto neste trabalho para esse fim é uma heurística baseada no K -means e no K -center, com livres adaptações para promover a obtenção de ambos os objetivos (máximas homogeneidade dos grupos e heterogeneidade dos componentes), gerando grupos de tamanhos iguais.

A ideia consiste em selecionar g indivíduos aleatórios, os quais serão os “centros” de cada grupo, e serão responsáveis por toda a geração dos mesmos. Os demais indivíduos terão calculadas suas distância a cada um dos centros, e serão alocados ao grupo associado ao centro mais distante que ainda tenha espaço de preenchimento disponível. Após a alocação de todos os indivíduos a um grupo, verifica-se a possibilidade de troca entre seus elementos de forma a promover um acréscimo na função objetivo (1). Com a realização de tais trocas, os agrupamentos obtidos são fornecidos como uma solução viável para o Problema (2) (Figura 1).



Figura 1: Esquema - Algoritmo proposto.

Considere uma população $U = \{P_1, \dots, P_m \mid P_i \in \mathbb{R}^n, i = 1, \dots, m\}$ de m indivíduos que possuem n características mapeadas. Como é desejada a formação de g grupos de tamanho t , temos que $m = tg$. No Passo 1 do Algoritmo 1, inicialmente, são selecionados aleatoriamente g centros, C_1, \dots, C_g , os quais serão os primeiros componentes dos grupos G_1, \dots, G_g . No Passo 2, cada indivíduo que ainda não pertence a um agrupamento é associado ao grupo cujo centro situa-se mais longe possível do mesmo, e que ainda não tenha atingido o tamanho máximo t . No Passo 3, é verificado se cada elemento não-centro s foi associado ao grupo cujo centro situa-se mais distante (grupo G_p), ou se o mesmo foi alocado em outro grupo (grupo G_q) pois G_p já estava totalmente preenchido. Se esse for o caso, a possibilidade de troca entre componentes de G_p e G_q é verificada no Algoritmo 2.

Algoritmo 1. Algoritmo Proposto:

Passo 0 - Entre com $U = \{P_1, \dots, P_m \in \mathbb{R}^n\}$, g e t .

Passo 1 - Para $i = 1, \dots, g$, faça:

Passo 1.1 - sorteie aleatoriamente $C_i \in U$;

Passo 1.2 - faça $U = U - \{C_i\}$, $\#G_i = 1$ e $G_i = \{C_i\}$.

Passo 2 - Para cada elemento s pertencente a U faça:

Passo 2.1 - calcule $q = \operatorname{argmax}_{\{j \mid \#G_j < t\}} \|s - C_j\|$;

Passo 2.2 - faça $\#G_q = \#G_q + 1$ e $G_q = G_q \cup \{s\}$.

Passo 3 - Para cada elemento s pertencente a U faça:

Passo 3.1 - calcule $q = \operatorname{argmax}_{\{j \mid \#G_j < t\}} \|s - C_j\|$ e $p = \operatorname{argmax}_j \|s - C_j\|$;

Passo 3.3 - se $q \neq p$, execute o algoritmo de análise de troca.

Observe que o algoritmo de análise de troca é executado somente se o indivíduo s , que atualmente pertence ao grupo G_q , tem maior preferência em participar do grupo G_p . Nesse caso, busca-se em G_p qual dos seus indivíduos está mais distante do centro de G_q (Passo 1). Denotando esse indivíduo por s_2 , no Passo 2 é realizada uma possível troca entre s e s_2 , gerando grupos modificados G'_p e G'_q . No Passo 3 são calculados os ganhos (ou perdas) da realização dessa troca para os grupos G_p e G_q , em termos da modificação produzida no valor final da função objetivo. Se, no geral, ocorrer um aumento na função ao executar essas duas modificações, elas serão realizadas e os grupos atualizados (Passo 4 - Figura 2).

Algoritmo 2. Análise de Troca:

Passo 0 - Entre com s, q e p .

Passo 1 - Calcule $s_2 = \operatorname{argmax}_{r \in G_p} \|r - C_q\|$.

Passo 2 - Faça $G'_p = G_p - \{s_2\} \cup \{s\}$ e $G'_q = G_q - \{s\} \cup \{s_2\}$.

Passo 3 - Calcule $\Delta_p = f(G'_p) - f(G_p)$ e $\Delta_q = f(G'_q) - f(G_q)$.

Passo 4 - Se $\Delta_p + \Delta_q > 0$, então faça $G_p = G'_p$ e $G_q = G'_q$.

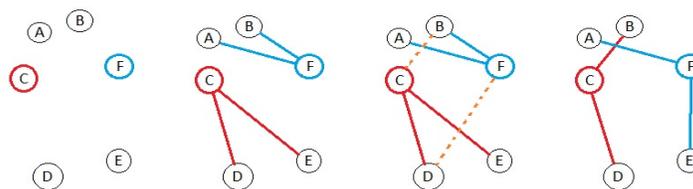


Figura 2: Exemplo - Etapas do agrupamento heterogêneo.

4 Resultados Numéricos

Neste trabalho, o algoritmo proposto foi utilizado para promover a formação de grupos em da disciplina Organização da Informação do curso de Bacharelado em Ciência da Computação da Universidade Federal do Rio de Janeiro - UFRJ, no segundo semestre de 2013. A turma composta de 60 alunos foi dividida em 15 grupos de 4 alunos, cada.

Para diferenciar um indivíduo do outro dentro do seu grupo utilizamos o perfil do estilo de aprendizagem. Entre os vários modelos de estilos de aprendizagem presentes na literatura de Psicologia Educacional [4], podemos destacar o de Felder-Silverman [6], que classifica os aprendizes em cinco dimensões: ativos/reflexivos; sensoriais/intuitivos; visuais/verbais; indutivos/dedutivos; seqüenciais/globais. A partir deste modelo, Felder e Soloman desenvolveram um instrumento denominado Índice de Estilos de Aprendizagem (*Index of Learning Styles - ILS*) [5], que classifica os estudantes em quatro das dimensões acima citadas. Cada aluno é, portanto, representado por um vetor de atributos de 4 componentes com valores ímpares variando de -11 a 11 .

Definido o perfil¹ através da identificação das características de seus estilos de aprendizagem, cada aluno é atribuído a um grupo de acordo com o Algoritmo 1 para a formação de agrupamento heterogêneo e também agrupamento homogêneo².

O resultado dessa divisão, representado na Tabela 1, evidencia, em valores, a heterogeneidade desejada. Representando a menor das mínimas diferenças entre os elementos de um mesmo grupo, $R(X)$, demonstra que, de fato, os elementos nos grupos heterogêneos estão mais distantes que nos grupos homogêneos. Bem como a maior distância entre os pontos médios, $M(X)$, evidencia que os grupos homogêneos tem a maior dessas diferenças. Em relação a um ponto médio dos elementos de cada grupo podemos concluir que a menor das médias das distâncias é, novamente, inferior nos grupos homogêneos. Assim como a média e a máxima distância em relação a esse ponto.

| Medidas | $R(X)$ | $M(X)$ | Distância média - grupos |
|-------------------------|--------|--------|--------------------------------------|
| Agrupamento Homogêneo | 2,00 | 17,95 | 2,97 (min); 5,09 (médio); 8,02 (max) |
| Agrupamento Heterogêneo | 2,82 | 4,82 | 5,26 (min); 7,49 (médio); 9,56 (max) |

Tabela 1: Resumo - Notas das avaliações.

Foi solicitado aos grupos que realizassem trabalhos que envolvessem pesquisa, organização, criação e apresentação, com as duas formas de agrupamento obtidas. Para o fim de avaliar o resultado dos grupos de estudos “podemos avaliar o produto final elaborado pelo grupo, ou ainda aquilo que cada componente aprendeu mediante sua elaboração. Também é possível propor uma auto-avaliação: que o grupo, ou cada integrante, se auto-avalie” [2]. Dessa forma, uma média das notas³ dadas pelo professor aos alunos mostra uma pequena diferença entre os grupos homogêneos, que tiveram uma média inferior a dos grupos heterogêneos nos trabalhos apresentados (Tabela 2).

| Avaliações | Média | Desvio Padrão | Nota mínima | Nota máxima |
|-------------------------|-------|---------------|-------------|-------------|
| Agrupamento Homogêneo | 7,6 | 3,10 | 0 | 10 |
| Agrupamento Heterogêneo | 9,1 | 0,47 | 8 | 9,5 |

Tabela 2: Resumo - Notas das avaliações.

¹Para maiores informações, consulte <http://www4.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/ILS.pdf>

²O agrupamento homogêneo também foi obtido por meio da execução do algoritmo proposto, mantendo a função objetivo e alterando o problema para minimização.

³Notas em escala de 0 à 10 encontradas no site da disciplina: <https://sites.google.com/site/orginfo201302/>

Os alunos também responderam um questionário com perguntas visando identificar a satisfação com sua participação no grupo, a participação de seus colegas e a satisfação com o resultado final do trabalho. Assim, nos grupos heterogêneos os alunos mostraram ter, na satisfação com seu trabalho e com o trabalho de seus colegas, uma vantagem de 5% em relação ao trabalho realizado em grupos homogêneos. A satisfação com o resultado final do trabalho foi equivalente em ambas formações de grupos.

5 Conclusões e Trabalhos Futuros

Os resultados apresentados pelo algoritmo proposto estão de acordo a formulação do problema. Estes implicam na heterogeneidade desejada entre os indivíduos do grupo, bem como os grupos gerados são homogêneos entre si. Assim, podemos dizer que a proposta atende aos objetivos formulados.

Há de ressaltar que quando lidamos com o fator humano, outros fatores além das características observadas podem influenciar no desempenho do aprendizado individual e no resultado de atividade do grupo. Por exemplo, a empatia entre as pessoas do grupo pode influenciar no resultado da satisfação de cada integrante. Por isso, são bem vindos novos estudos que abordem outros tipos de características para agregar mais heterogeneidade ao grupo.

Além de grupos de estudos, a importância dos agrupamentos heterogêneos são evidentes nas pesquisas de Kaoro Ishikawa onde, em uma das etapas, são selecionados indivíduos de níveis hierárquicos e setores diferentes para a solução de problemas de diversas áreas das empresas. Tal prática é apresentada na teoria sobre os Círculos de Controle de Qualidade (CCQ) [3]. O algoritmo, aqui proposto, pode auxiliar a aplicação da técnica CCQ.

Referências

- [1] S. E. Asch, “Psicologia Social”, Editora Nacional, 1977, 4ed, São Paulo.
- [2] J. Bonals, “O trabalho em pequenos grupos na sala de aula”, Artmed, 2003.
- [3] V. F. Campos, “TQC: Controle da Qualidade Total (no estilo japonês)”, Editora de Desenvolvimento Gerencial, Belo Horizonte-MG, 1999.
- [4] R. M. Felder and E. M. Henriques, “Learning and teaching styles in foreign and second language education”, *Foreign Language Annals*, vol.28, 21-31, 1995.
- [5] R. M. Felder and Joni Spurlin, “Applications, Reliability and Validity of the Index of Learning Styles”, Tempus Publications, 2005.
- [6] R. M. Felder, “Matters of Style”, *ASEE Prism*, 6(4), 18-23, 1996.
- [7] A. D. Gordon, “Classification”, Chapman and Hall, 1981.
- [8] J. Han and M. Kamber, “Cluster Analysis”, Morgan Kaufmann Publishers, 2001.
- [9] K. A. Jehn and G. B. Northcraft and M. A. Neale, “Why differences make a difference: a field study of diversity, conflict and performance in workgroups.”, *Administrative Science Quarterly*, 44, december, 741-763, 1999.
- [10] D. I. Jung and J. J. Sosik and K. B. Baik, “Investigating work group characteristics and performance over time: a replication and cross-cultural extension.”, *Group Dynamics Theory: Research and Practise*, 6 (2), 153-171, 2002.
- [11] L. Kaufman and P. J. Rousseeuw, “Finding Groups in Data: An Introduction to Cluster Analysis (Wiley Series in Probability and Statistics)”, WileyInterscience, 2005.

- [12] K. Lewin, “Problemas de dinâmica de grupo”, Cultrix, 1948, São Paulo.
- [13] J. B. MacQueen, “Some methods for classification and analysis os multivariate observations”, Fifth Berkeley Symposium on Mathematical Statistic ans Probability, 1:281-297,1967.
- [14] A. N. Melser, “Gifted students and cooperative learning: A study of grouping strategies.”, Roper Rev,315, 1999.
- [15] G. M. Parker, “O Poder das Equipes: um guia prático para implementar equipes interfuncionais de alto desempenho”, Campus, 1995.
- [16] M. E. Shaw, “Group dynamics: the psychology of small group behavior”. 3 ed. EUA: McGraw-Hill, 1981.
- [17] P. Tan, M. Steinbach, V. Kumar, “Introducing to Data Mining”, Addison-Wesley Longman Publishing Co., 2005.
- [18] H. E. Triandis and E. R. Hall and R.B. Ewen, “Member heterogeneity and dyadic creativity”, Human Relations, 1965.
- [19] W. E. Watson and K. Kumar and L. K. Michaelsen, “Cultural diversity´s impact on interaction process and performance: comparing homogeneous and diverse task groups.”, Academy of Management Journal, 36 (3), june, 590-598, 1993.