Formulações Matemáticas e Avaliação de Resolução do *Layout* de Equipes de Desenvolvimento em Fábricas de Software

João Amilcar Rodrigues, Ítalo Yeltsin, Bruno Monteiro, Juan Garcia, Marcos Negreiros, Pablo Batista, Augusto Palhano
Universidade Estadual do Ceará (UECE)
Mestrado Profissional em Computação Aplicada - MPCOMP/UECE-IFCE
Mestrado Acadêmico em Ciência da Computação (MACC-UECE)
Programa de Pós-Graduação em Administração (PPGA-UECE)
Av. Silas Munguba, 1700 – Campus do Itaperi
CEP: 60740-000 – Fortaleza – CE – Brazil
joaoamilcar@gmail.com, br.yeltsin@gmail.com,
brunex92@gmail.com, juan.garcia.cc2011@gmail.com,
negreiro@graphvs.com.br, opablofernandes@hotmail.com,
apalhano@graphvs.com.br

RESUMO

Este trabalho apresenta um estudo do processo de modelagem e aplicação de modelos matemáticos do Problema de Agrupamento Capacitado, ao problema de "layout" de times de desenvolvimento de projetos em fábricas de software. Foi realizada a investigação de uma solução abordando um problema de disposição de times de projetos de TI, indicado pela fábrica de software da DataPrev, Fortaleza/CE. Os modelos abordados consideram uma variação generalizada do problema de p-Medianas Capacitadas (PpMC), problema de agrupamento capacitado com o menor maior diâmetro (PCCMMD) e o problema de agrupamento capacitado heterogêneo com centro geométrico (PACHCG). São apresentadas e avaliadas as soluções obtidas usando os *solvers* mais adequados disponíveis na plataforma Argonne Labs – NEOS Server.

PALAVRAS CHAVE. Agrupamento, Modelagem, “Solvers".

Tópicos: Aplicações de PO, Programação Inteira e Combinatória

ABSTRACT

This work presents a study of modeling and application process of mathematical models for Capacitated Clustering Problem, applied to placement (layout) of development team in software companies. A development project and its solution was elaborated to reach a real problem in the IT Software Factory team placement indicated by DataPrev, Fortaleza/CE. The discussed models are: a generalized variation of the Capacitated p-Medians problem (CpMP); the minimum maximum diameter capacitated clustering problem (MMDCP) and capacitated heterogeneous clustering problem with geometric center (GCHCCP). Here are also presented and evaluated the obtained solutions using the most appropriate solvers available in Argonne Labs - NEOS Server platform.

KEYWORDS. Clustering, Modeling Process, MP Solvers.

Paper topics: OR Application, Combinatorial Optimization

1. Introdução

A dificuldade de se resolver problemas combinatórios reais é eminente e os desafios sobre a compreensão da realidade e sobre o que se deve de fato levar em consideração sobre ela, é um problema que normalmente se depara os especialistas em otimização durante o processo de concepção de modelos matemáticos. Nos foi colocado um desafio pelo Centro de Desenvolvimento de Sistemas da DataPrev, Fortaleza/CE, sugerindo-se resolver o problema de disposição ("layout") das equipes de desenvolvimento de sistemas computacionais da fábrica de software de Fortaleza.

As equipes de desenvolvimento da DataPrev se organizam em dois grandes salões de um prédio, em andares distintos (3o e 8o), onde estão dispostas as estações de trabalho em células de até quatro estações, Figura 1. São ao todo 175 lugares onde as equipes são alocadas. Cada equipe tem um número de pessoas a ela associado já previamente definido, sendo este número normalmente distinto entre equipes em razão das diferentes complexidades de projetos que desenvolverão.

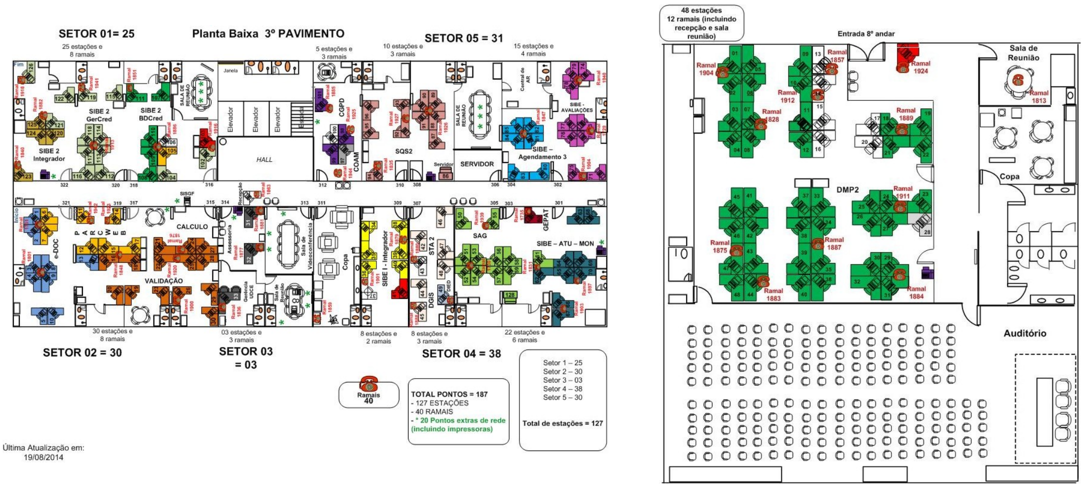

Figure 1: Disposição das estações de trabalho na DataPrev, Fortaleza/CE.

Duas pessoas são responsáveis pela distribuição das equipes nos dois andares. Conforme os projetos vão concluindo, é sempre um grande problema para elas realocar as equipes em função das estações disponíveis. Apesar disso, o problema básico seria resolver a alocação das equipes conforme a melhor disposição nas estações disponíveis. As gerentes da fábrica tem grandes problemas para fazer esta disposição, em razão das equipes de projeto serem muitas e terem diversos tamanhos.

Modelos matemáticos para atender a este problema ainda não foram propostos pela literatura, até onde sabemos. Desse modo, consideramos então pesquisar abordagens semelhantes ou mesmo adaptações de modelos combinatórios já existentes, de modo a definir o melhor modo de solução para o problema desafio proposto.

Apresentamos a seguir na seção 2 as descrições dos modelos utilizados para os problemas, na seção 3 apresentamos os resultados obtidos com a aplicação dos modelos com os *solvers*, e as análises referentes a cada aplicação. Por fim, na seção 4 mostramos a direção referente ao melhor modelo de decisão para a empresa.

2. Modelos Matemáticos Adotados

Dentre os problemas de programação matemática mais próximos ao proposto podemos considerar: o problema das p-medianas capacitadas, o problema do menor-maior diâmetro em agrupamentos capacitados, e o problema de agrupamento capacitado em centro geométrico. Tais problemas foram estudados pela literatura em contextos diferentes do aqui apresentado, necessitando obviamente ajustes e revisão conceitual para o caso da Dataprev. As revisões conceituais permitem uma mudança tal que novos modelos precisam ser construídos, não simplesmente ajustados, como veremos a seguir.

2.1. O Problema das p-Medianas Capacitadas

Este problema considera que são dados itens de demanda/oferta conhecida, e sua respectiva posição no espaço. Deseja-se encontrar, a partir de um conjunto de medianas com capacidade limitada e posição conhecida, a atribuição completa que minimiza a distância total itens-mediana. Neste caso, e relacionado ao problema proposto pela DataPrev, as medianas seriam as estações de trabalho onde ficariam os gerentes ou líderes das equipes, enquanto os itens seriam as estações onde ficariam as demais pessoas de sua equipe.

No modelo de p-medianas capacitadas introduzido por Mulvey & Beck (1984) considera-se que o conjunto das medianas e suas capacidades são previamente conhecidas, Mulvey e Beck [1984]. Sem fazer qualquer modificação neste modelo seria inviável ajustá-lo ao problema da DataPrev. Como no problema real, precisamos definir os grupos e medianas a partir das estações de trabalho. A melhor alternativa encontrada foi refazer o modelo de Mulvey & Beck agora considerando que qualquer estação poderia ser uma mediana.

A formulação do problema Generalizado de p-Medianas Capacitadas (PgpMC), segue a seguinte descrição:

Conjuntos:

I - é o conjunto dos itens (ou estações);

K - é o conjunto de grupos (ou equipes);

$|K|$ - é a cardinalidade do conjunto K, ou um número fixo de grupos ($|K| = p$);

Parâmetros:

d_{ij} - é a distância entre uma estação i e a sua estação mediana j;

Q_k - é a capacitade máxima do grupo k;

q_i - é a demanda da estação i (q_i = 1, para todas as estações em nosso caso);

Variáveis:

$$g_k = \begin{cases} 1, \text{se o grupo } k \text{ é usado do conjunto de grupos possíveis;} \\ 0, \text{caso contrário} \end{cases}$$

$$x_{ij} = \begin{cases} 1, \text{se o item } i \text{ está alocado à mediana } j; \\ 0, \text{caso contrário} \end{cases}$$

O problema gpMC pode ser formulado como:

$$(\mathbf{gpMC}) Minimizar \sum_{i \in I} \sum_{j \in I} d_{ij}\, x_{ij} \tag{1}$$

$$\text{tal que}: \sum_{j \in I} x_{ij} = 1, \forall i \in I \tag{2}$$

$$x_{ij} \leq x_{jj}, \forall i \in I, \forall j \in I, \tag{3}$$

$$\sum_{i \in I} x_{ii} = \sum_{k \in K} g_k, \tag{4}$$

$$\sum_{i \in I} q_i\, x_{ij} \leq Q_k\, g_k, \forall j \in J \tag{5}$$

$$\sum_{j \in I} x_{jj} = p, \tag{6}$$

$$g_k \in \{0,1\}, \forall k \in K \tag{7}$$

$$x_{ij} \in \{0,1\},\ \forall i \in I, \forall j \in J \tag{8}$$

A função objetivo 1 minimiza a distância total entre as medianas e cada item atribuído a sua mediana. A restrição 2 indica que o item deve ser atribuído a apenas uma mediana. A restrição 3 define que uma vez uma mediana seja usada um item deve ser atribuído a ela. A restrição 4 indica que o número de medianas usadas deve ser o mesmo que o número de medianas abertas. A restrição 5 considera que a soma das demandas dos itens não ultrapassa a capacidade da mediana a que eles são atribuídos. A restrição 6 limita o número de medianas usadas. As restrições 7 e 8 se referem às variáveis de decisão do modelo.

O problema gpMP é NP-Hard assim como sua versão original, Mulvey e Beck [1984].

2.2. O Problema do Menor-Maior Diâmetro de Agrupamento Capacitado

Nesta versão do Problema de Agrupamento Capacitado, deseja-se minimizar entre os grupos a serem formados aquele que possui o maior diâmetro, respeitando-se o limite de capacidade de cada grupo a ser aberto. Este problema foi extraído de uma versão irrestrita do problema de agrupamento de menor-maior diâmetro proposto por Roi [1971]. Aqui foi inserida a restrição de capacidade por grupo para atender ao nosso propósito específico.

Considerando o problema de *layout* que tratamos, ao usarmos este modelo estamos apenas preocupados com a formação dos grupos. Sendo uma formulação mais simples que a anterior, imaginamos obter vantagens de tempo e possibilidade de obtenção de uma boa solução na resolução do problema real.

A formulação do Problema de Menor-Maior Diâmetro de Agrupamento Capacitado (PM-MDAC), pode ser assim descrita:

Conjuntos:

I - é o conjunto de estações ($|I| = n$);

M - é o conjunto de grupos ($|M| = m$);

Parameters:

p - é o número de grupos;

d_{ij} - é a distância da estação i à estação j;

q_i - é a demanda de uma estação i;

Q_j - é a capacidade máxima do grupo j;

Variáveis:

Z - é o maior diâmetro entre os m grupos;

$$x_{ik} = \begin{cases} 1, \text{Se uma estação } i \text{ está atribuída ao grupo } k; \\ 0, \text{caso contrário} \end{cases}$$

O problema MMDAC pode ser formulado como:

$$(\mathbf{MMDAC}) Minimizar \ \ Z \tag{9}$$
$$\text{tal que}: \tag{10}$$
$$d_{ij}x_{ik} + d_{ij}x_{jk} - Z \leq d_{ij}, \ \ \forall i \in I, \forall j \in I, \ i \neq j, \ \forall k \in M \tag{11}$$
$$\sum_{k \in M} x_{ik} = 1, \ \ \forall i \in I \tag{12}$$
$$\sum_{i \in I} q_i \, x_{ik} \leq Q_k, \ \ \forall k \in M \tag{13}$$
$$Z \geq 0, \ x_{ik} \in \{0,1\}, \ \forall i \in I, \forall k \in M \tag{14}$$

A função objetivo 9 minimiza o diâmetro do grupo de maior diâmetro. As restrições 11 indicam que as estações i e j devem ser atribuídas ao grupo k se a distância entre elas não ultrapassa o maior diâmetro Z dos grupos abertos. As restrições 12 atribuem uma estação a um grupo. As restrições 13 consideram que a soma das estações atribuídas a um grupo não pode ultrapassar a capacidade deste grupo. As restrições 14 se referem às variáveis de decisão do modelo.

O MMDAC também é um problema NP-HARD, assim como a sua versão irrestrita.

2.3. O Problema de Agrupamento Capacitado Heterogêneo com Centro Geométrico

O problema original proposto por Negreiros & Palhano (2006) a capacidade dos grupos é homogênea, e a função objetivo considera a minimização da variância total entre indivíduos e o centros dos grupos a que pertencem. Aqui consideramos que é necessário formar grupos de diferentes capacidades e minimizar a distância total entre as estações e os centros dos grupos a que elas são atribuídas. A solução deste modelo é particularmente interessante para o nosso problema real, porque retorna além da disposição dos grupos, uma posição próxima aonde o gerente ou líder de cada equipe deve ficar.

A formulação do Problema de Agrupamento Capacitado Heterogêneoo em Centro Geométrico (pACHCG), pode ser assim descrita:

Conjuntos:

r - é a dimensão do espaço Euclideano ($r = 2$, em nosso caso);

I - é o conjunto das estações;

J - é o conjunto de centros dos grupos;

$|J|$ - é a cardinalidade do conjunto J, ou um número fixo de grupos ($|J| = p$);

Parâmetros:

p - é o número de grupos;

a_i - é um vetor de dimensão r com as coordenadas da estação i;

q_i - é a demanda de uma estação i;

Q_j - é a máxima capacidade de um grupo j;

n_j - é o número de indivíduos no grupo j;

Variáveis:

$$\bar{x}_j = \begin{cases} 1, \text{é um vetor de dimensão } r \text{ representando o centro das coordenadas do } \textit{grupo } j \\ 0, \text{caso contrário} \end{cases}$$

$$y_{ij} = \begin{cases} 1, \text{Se uma estação } i \text{ está atribuída a um grupo } j \\ 0, \text{caso contrário} \end{cases}$$

O Problema pACHCG pode ser formulado como:

$$(\mathbf{pACHCG}) Minimizar \sum_{i \in I} \sum_{j \in J} ||a_i - \bar{x}_j|| \, y_{ij} \tag{15}$$

$$\text{tal que} : \sum_{j \in J} y_{ij} = 1, \forall i \in I \tag{16}$$

$$\sum_{i \in I} y_{ij} \leq n_j, \forall j \in J \tag{17}$$

$$\sum_{i \in I} a_i \, y_{ij} \leq n_j \bar{x}_j, \forall j \in J \tag{18}$$

$$\sum_{i \in I} q_i \, y_{ij} \leq Q_j, \forall j \in J \tag{19}$$

$$\bar{x}_j \in \Re^r, n_j \in \aleph, \forall j \in J \tag{20}$$

$$y_{ij} \in \{0, 1\}, \; \forall i \in I, \forall j \in J \tag{21}$$

A função objetivo 15 minimiza a distância Euclideana total entre os centros dos *grupos* e estações a eles atribuídas. A restrição 16 atribui um indivíduo a apenas um *grupo*. A restrição 17 considera o número de estações por *grupo*. A restrição 18 define o centro geométrico dos *grupos*. A restrição 19 limita o número de estações atribuídas a um grupo à sua máxima capacidade correspondente Q_j. As restrições 20 e 21 se referem às variáveis de decisão do modelo.

O modelo pACHCG é não-linear e binário, sendo o problema NP-Hard assim como suas variações, Hansen e Jaumard [1997], Negreiros e Palhano [2006], Stefanello e Muller [2009], Batista et al. [2014], Prata [2015].

3. Resultados Alcançados

Os modelos foram executados utilizando os seguintes *solvers*: Gurobi, Xpress, Mosek e Filmint disponíveis na plataforma ARGONNE NEOS SERVER, Neos [2015]. Os dois primeiros resolvem apenas problemas mistos binários-inteiros e lineares, enquanto os demais resolvem problemas não-lineares binários-inteiros, Abhishek et al. [2008]. Em conjunto, estes *solvers* são indicados à resolução dos dois tipos de modelos que foram usados nas nossas avaliações.

As instâncias provenientes da base real da DataPrev, com 175 estações foram avaliadas para os casos de capacidade heterogênea e homogênea. Nesta instância, para que pudessemos tratar a presença do 8o andar, colocamos o conjunto de estações no mesmo plano, porém bastante afastadas, de modo a nos ajudar quanto à coerência da solução de *layout*, figura 2.

Nas tabelas a seguir, *solver* indica o *software de otimização* usado para resolver o modelo matemático pertinente, *FO* indica o valor da função objetivo final obtido pelo *solver* na instância, e *Tempo (s)* é o tempo em segundos que o *solver* levou para resolver a instância do problema.

Na tabela 1 apresentamos os resultados para a instância DataPrev (real), com capacidades de grupo distintas (heterogênea). Nota-se que para o PACMMD os *solver* Gorubi e Xpress mantem a concordância quanto à solução, porém o software Xpress é 3,6 vezes mais rápido. Para este caso as instâncias do PgpMC não foram resolvidas por ambos os *solvers*. As figuras 2 e 3 apresentam uma compilação no plano da instância DataPrev e a melhor solução obtida usando o solver Filmint, como é mostrado na tabela 1.

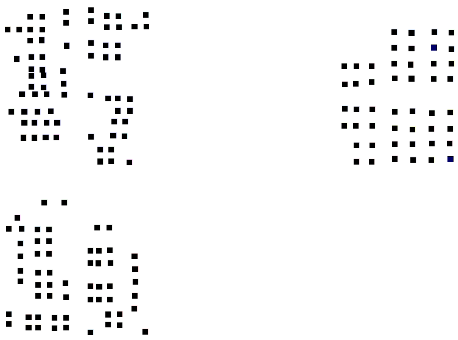

Figure 2: Instância DataPrev ajustada para ser avaliada.

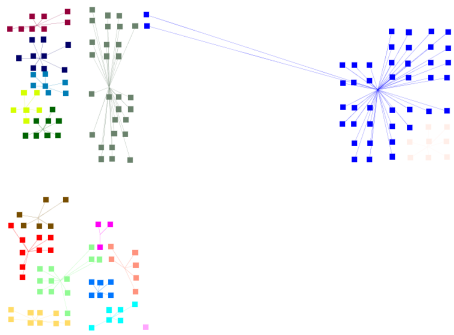

Figure 3: Solução do *solver* FILMINT para a versão heterogênea capacitada.

A solução obtida pelo *solver* Filmint para o PACHCG, foi indicada como ótima. Por fim, na solução gerada pelo software Xpress para o MMDAC, nos mostrou que os grupos se misturaram no 3o andar devido ao diâmetro do maior grupo, que possui estações nos dois andares.

Instância DataPrev n=175, $Q_{1...17}$=[42, 9, 6, 1, 9, 8, 9, 8, 7, 3, 8, 6, 12 , 6, 6, 6, 29]				
Modelo Matemático	Solver		Solver	
	Tempo(s)	FO	Tempo(s)	FO
PAMMD	Gorubi		Xpress	
	302.426	795.083	83.7613	878.492
pACHCG	Filmint			
	0.120981	35440.3		
gpMC	Gorubi		Xpress	
	-		-	

Table 1: Resultados para os três modelos sobre instância DataPrev (real) heterogênea.

Na tabela 2 apresentamos os resultados para a instância artificial com capacidades de grupo iguais (homogênea). Nota-se que para o PAMMD os *solvers* Gorubi e Xpress concordam quanto à solução, sendo o Xpress duas vezes mais rápido. Já para o modelo do PgpMC o solver Gorubi retornou a melhor solução com *gap* de 3,63% sobre o *solver* XPress. Para este caso as resoluções da instância do PgpMC foram indicadas serem resolvidas na otimalidade para ambos os *solvers* Gorubi e Xpress. Nota-se também as sutis diferenças no formato final da solução. Além disto para ambas as soluções há um grupo no 8o andar que se une a uma estação no pavimento mais baixo, o que não é desejável. As figuras 4 e 5 mostram essa composição.

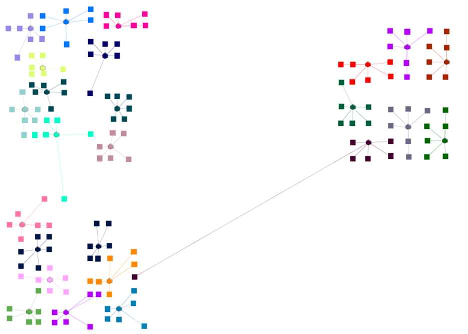

Figure 4: Solução capacitada hemomogênea em p-Mediana Capacitada obtida pelo *solver* XPress.

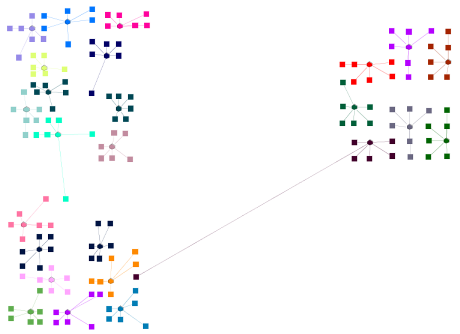

Figure 5: Solução capacitada hemomogênea em p-Mediana Capacitada obtida pelo *solver* Gorubi.

Conforme a tabela 2, na solução obtida pelo software FilMint para o PACHCG, o *solver* indicou a prova de sua otimalidade. Para o problema gpMC, os *solvers* resolveram a instância do problema discordando da solução, tendo o *solver* Xpress resolvido 200 vezes mais rápido, porém obtendo um resultado pior de custo total.

Instância DataPrev n=175, $Q_{1...25}$=7				
Modelo Matemático	Solver		Solver	
	Tempo(s)	FO	Tempo(s)	FO
PAMMD	Gorubi		Xpress	
	807.13	1398.01	414.966	1398.01
pACHCG	FilMint			
	0.119981	21409.7		
gpMC	Gorubi		Xpress	
	18994.8	19477.9	91.2691	19634.6

Table 2: Resultados para os três modelos sobre a instância DataPrev homogênea.

Para esse caso, a solução que retorna o modelo homogêneo min-max capacitado obtida por ambos os *solvers* Xpress e Gorubi é de baixa qualidade, não ajudando no processo, figura 6. Já curiosamente para o modelo de centro geométrico capacitado obtida pelo *solver* Filmint, a solução não gera união de grupos entre os dois pavimentos, mantendo uma compactação muito mais favorável que aquela apresentada pelo resultado do modelo com p-medianas capacitadas.

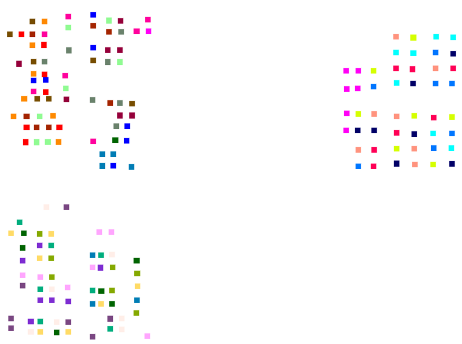

Figure 6: Solução homogênea em Agrupamento Min-Max Capacitado obtida pelo *solver* Xpress.

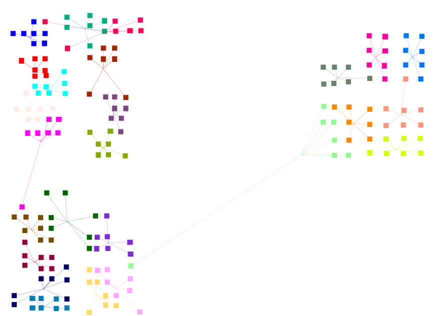

Figure 7: Solução homogênea em Agrupamento Capacitado obtida pelo *solver* Filmint.

Pela figura 7, curiosamente é fácil ver que se encontrarmos a mediana de cada grupo resultante do modelo em agrupamento capacitado em centro geométrico, a solução final será bem inferior às obtidas pelos *solvers* para os casos em PgpMC. Isto sugere obviamente que as soluções obtidas naqueles *solvers* jamais poderiam ser ótimas.

4. Conclusões

Este trabalho apresentou um projeto de modelagem matemática e resolução para um problema de otimização do *layout* e times de desenvolvimento de sistemas computacionais da fábrica de software da DataPrev. Dentre os modelos abordados, discutimos a sua formulação, o ajuste a partir de outros modelos já conhecidos e mesmo a remodelagem. Aplicamos aos modelos os principais *solvers* do mercado para resolvê-los a partir da instância fornecida pela empresa. Foram verificadas instâncias reais e artificialmente geradas para calibrar e verificar os modelos. O Modelo pACHCG é o modelo que permite conseguir achar solução viável para as instâncias reais e artificiais, heterogêneas ou homogêneas avaliadas

O desafio apresentado foi interessante, pois além do estudo de modelos apropriados para o problema, verificamos também que os *solvers* discordaram em soluções de modelos lineares binários, indicando a assinatura de otimalidade para os casos investigados. Notamos também que seria possível obter através de um modelo não linear binário (PACHCG) solução muito mais atraente e de menor custo que as apresentadas anteriormente, se ajustadas ao modelo de medianas.

Para a instância proposta pela DataPrev, o modelo mais adequado de ser usado com os *solvers* disponíveis na NEOS Argonne, foi o modelo para o Agrupamento Capacitado Heterogêneo com Centro Geométrico (PACHCG), resolvido pelo *solver* Filmint. Apesar de ser o escolhido, a solução obtida juntou no mesmo grupo estações dos dois pavimentos, algo *a priori* indesejável. Apesar disto, seria ainda necessário ajustar o modelo para não permitir a colocação de turmas em diferentes pavimentos, que na verdade pode ser facilmente atendida pela inclusão de um custo elevado de distância entre os indivíduos de pavimentos distintos diretamente na função objetivo de modo a evitar este tipo de junção.

References

Abhishek, K., Leyffer, S., e Linderoth, J. T. (2008). *FilMINT: An Outer-Approximation-Based Solver for Nonlinear Mixed Integer Programs*, volume ANL/MCS-P1374-0906. ARGONNE NATIONAL LABORATORY.

Batista, P., Negreiros, M., e Palhano, A. (2014). Solvers resolution perspective to the capacitated centred clustering problem. In *Anais do XLVI SBPO*, p. 2458–2468. SOBRAPO.

Hansen, P. e Jaumard, B. (1997). Cluster analysis and mathematical programming. *Mathematical Programming*, p. 191–215.

Mulvey, J. e Beck, M. (1984). Solving capacitated clustering problems. *European Journal of Operational Research*, 18:339–348.

Negreiros, M. e Palhano, A. (2006). The capacitated centred clustering problem. *Computers and Operations Research*, 33:1639–1663.

Neos (2015). Neos (network-enabled optimization system). `http://www.neos-server.org/neos/solvers/index.html`.

Prata, B. (2015). The multi capacitated clustering problem. Technical report, Universidade Federal do Ceará.

Roi, M. (1971). *Cluster analysis and mathematical programming*, volume 66. Journal of the American Statistical Association.

Stefanello, F. e Muller, F. (2009). Um estudo sobre problemas de agrupamento capacitado. In *Anais do XLI SBPO*, p. 2819–2828. SOBRAPO.