

Algoritmos Genéticos na obtenção de uma Grade de Horários com Múltiplos Cursos para uma Instituição de Ensino

Alexandre Brasil da Silva¹, Carlos Michel Betemps¹, Milton Heinen¹

¹Universidade Federal do Pampa (UNIPAMPA) – Bagé – RS – Brasil

alexandre.brasil.eng.comp@gmail.com, {carlos.betemps,
miltonheinen}@unipampa.edu.br

Abstract. *This article presents the application of a heuristic technique – genetic algorithms – solving the problem in the class schedule for multiple courses in an educational institution. The goal is to sweep the solution space through a guided technique, so that the timetable concerning the resources can be allocated efficiently and offer the possibility of real application. The technique was shown to be applicable in practice through the use of bonus and penalty criteria of the solutions, and was suitable for the hourly grids generated results.*

Resumo. *Este artigo apresenta a aplicação de uma técnica heurística – algoritmos genéticos – na resolução do problema da grade horária para múltiplos cursos em uma instituição de ensino. O objetivo foi varrer o espaço de solução através de uma técnica guiada, para que os recursos concernentes à grade horária possam ser alocados de maneira eficiente e que ofereçam possibilidade de aplicação real. A técnica mostrou-se aplicável na prática por meio do uso de critérios de bonificação e penalização das soluções, sendo que apresentou resultados adequados para as grades horárias geradas.*

1. INTRODUÇÃO

O problema de confecção da grade horária é comum à grande maioria das instituições de ensino [Ferreira e Corrêa 2010] e a sua solução é denotada na distribuição sistemática de encontros entre docentes e discentes durante o período letivo [Gervásio 2012]. Essa distribuição deve levar em consideração alguns fatores/restrições como o de professores ministrarem disciplinas com as quais possuem afinidades, docentes não serem alocados em duas salas num mesmo horário concomitantemente, uma mesma sala não ser ocupada por duas turmas com disciplinas diferentes num mesmo período [Simonetti 2007], etc.

É visto que quanto maior o número de recursos (docentes, salas, disciplinas, turmas), maior será o número de combinações que poderão ser feitas, tornando mais fácil a ocorrência de erros e colisões de disciplinas e docentes. Se a solução for buscada de forma manual acaba tornando-se inviável, necessitando de um fator computacional como meio auxiliar nesta atividade [Lobo 2005].

Deste modo, este problema ainda pode possuir um conjunto de soluções que satisfazem as restrições de seu escopo, formando um conjunto de respostas corretas. Ele é um dos clássicos problemas da computação [De Oliveira 2000], cujos autores denominam *timetabling problem*, e é classificado como pertencente à classe dos problemas NP-Difíceis [Gervásio 2012].

A inteligência artificial, através de suas técnicas de busca, tem sido muito empregada em problemas que possuem um elevado número de recursos que devem ser

combinados, alocados, e submetidos a determinadas restrições. Entre estas técnicas, destacam-se os algoritmos genéticos, que muitas vezes apresentam resultados extremamente satisfatórios para estes tipos de problemas [Lobo 2005].

Este tema já foi abordado em vários trabalhos, como em [Ferreira e Corrêa 2010], [Gervásio 2012], [Simonetti 2007], [Lobo 2005], [De Oliveira 2000]. O que os difere é o método como os genes, os indivíduos e os operadores genéticos são implementados. São utilizadas as mais diversas estruturas de dados e inúmeras variações do fluxograma clássico de um algoritmo genético.

Neste artigo, as seções ficam divididas em: Seção 2 (Metodologia e Desenvolvimento do Trabalho) que explana os componentes e como o problema foi modelado para ser tratado pela técnica em questão. Seção 3 (Experimentos e Resultados) mostra quais resultados foram obtidos a partir dos testes realizados e a Seção 4 (Conclusões) expõe as conclusões obtidas até o presente momento.

2. METODOLOGIA E DESENVOLVIMENTO DO TRABALHO

Os algoritmos genéticos, dentro da inteligência artificial, são uma subárea dos algoritmos evolucionários. Estes são altamente inspirados na teoria da seleção natural de Charles Darwin, em que indivíduos mais aptos sobrevivem por maior tempo, e têm, desta forma, maior chance de realizarem cruzamentos, fazendo com que eles possam carregar seu material genético mais vezes às gerações futuras. Desta maneira, o algoritmo pode direcionar sua pesquisa dentro do espaço de buscas através da realização de um tipo de “refinamento da espécie”. Esta é uma técnica aleatório-guiada, em que leva-se em consideração as iterações ocorridas para tornar mais eficiente sua passagem pelo espaço de buscas (guiada), porém, sempre levando em consideração, informações que estão dentro de um universo probabilístico (aleatório)[Linden 2012].

Conforme [De Oliveira 2000], [Linden 2012] e [Lobo 2005], a escolha dos Algoritmos Genéticos para o problema em questão, deu-se, entre outros fatores, principalmente por: poderem resolver problemas complexos de forma confiável; a construção de um Algoritmo Genético e de seus modelos é algo geralmente simples de ser implementado; são extensíveis a outros problemas; existe uma certa facilidade de combinação com outros métodos de busca, principalmente no que se refere a questão de otimização dos seus parâmetros de execução; existe a plena possibilidade de implementação que tire vantagem de arquiteturas paralelas.

As informações sobre o problema, e como elas serão modeladas dentro dos genes/indivíduos, não seguem uma regra, elas somente devem representar o problema da melhor, e mais eficiente, forma possível [Linden 2012].

Enumeram-se abaixo os elementos necessários no AG implementado:

Gene: parcela indivisível de um indivíduo, implementado na forma de uma tupla com as informações de disciplina, docente, sala, semestre e curso.

Indivíduo: é formado por um conjunto de genes e é uma das soluções candidatas à resposta para o problema. Foi implementado na forma de uma lista de matrizes, onde cada posição de suas matrizes é ocupada por um gene. Cada objeto desta lista refere-se à grade horária de determinado semestre, e a matriz da grade do último semestre de determinado curso é seguida pela matriz que possui a grade horária do primeiro semestre do próximo curso selecionado para confecção dos horários.

Como pode ser observado na figura 2, as 4 primeiras linhas de cada matriz representam os períodos da manhã e da tarde, e as duas últimas linhas representam os

períodos noturnos. A linha identificada por “Intervalo Turnos” (quinta linha) é o período de intervalo entre os turnos (noturno/diurno) e pode ser usada como período letivo. Cada linha nesta representação equivale a dois períodos consecutivos (forma usual de alocação na UNIPAMPA Campus Bagé) onde cada gene terá na verdade, dois períodos consecutivos de uma mesma disciplina.

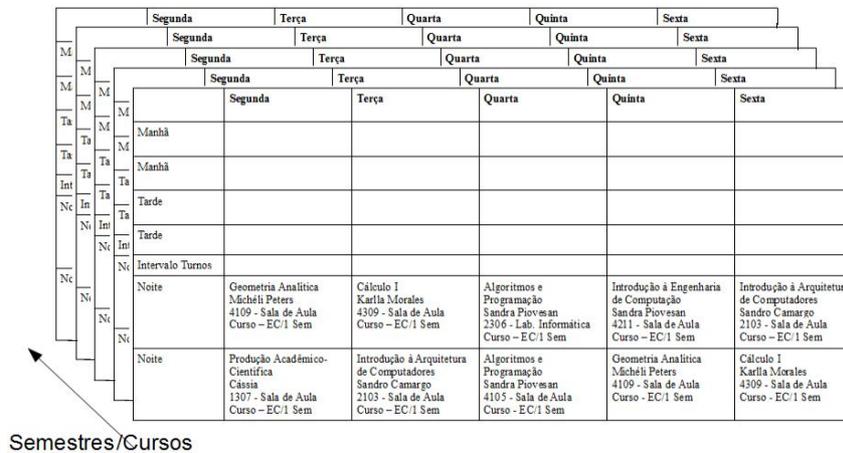


Figura 10. Indivíduo implementado, com suas matrizes.

População: conjunto de n indivíduos (parametrizado antes da execução do algoritmo) coexistindo no período t de tempo. Na primeira geração (população inicial), este t é zero.

População Inicial: Criada de forma aleatória, apenas combinando nos genes as informações de disciplinas, docentes, salas, semestres e cursos, e distribuindo-os dentro das matrizes dos indivíduos. Os únicos itens observados são os de não alocar disciplinas à docentes que não possuem afinidade com a mesma, e de não alocar disciplinas em semestres que são diferentes dos semestres/cursos nos quais são ofertadas.

Geração: variação dos indivíduos da população conforme as gerações (t) aumentam. É um valor maior ou igual a 0; acrescido de 1 a cada iteração do algoritmo.

Função de Avaliação: função que calcula a aptidão (métrica) de um indivíduo em relação à resposta esperada para o problema. Devem ser levadas em consideração, restrições a serem obrigatoriamente atendidas: a não ocorrências de aulas concomitantes para um mesmo docente em salas diferentes e não haver disciplinas diferentes alocadas em uma mesma sala em um mesmo horário. Existem também restrições que preferencialmente devem ser satisfeitas (mas caso não sejam satisfeitas, não invalidam a resposta): alocação dos docentes conforme suas preferências de horários (indisponíveis, disponíveis e preferenciais) e o controle da carga horária distribuída entre os docentes.

Cada gene alocado em uma matriz é um bônus para o indivíduo, assim como quando um docente é alocado em um período em que o mesmo tem preferência por ministrar suas aulas. Quando um docente é designado em um período em que é tido como indisponível, quando é alocado em duas salas num mesmo período e quando disciplinas diferentes são alocadas simultaneamente numa mesma sala, são penalizações.

Também é tido como penalização, o resultado do cálculo do desvio padrão na distribuição da carga horária entre os docentes. Indivíduos que possuem carga horária distribuída de maneira mais igualitária (menor desvio padrão) são menos penalizados nas suas avaliações do que indivíduos em que o desvio padrão é mais alto.

A equação 1 mostra a função que calcula a aptidão de um indivíduo.

$$f(x) = \sum \text{bônus} - \sum \text{penalizações}$$
 Equação (1)

Método de Seleção: É como os indivíduos são selecionados para que possam realizar os cruzamentos entre si ou mutações originando novos indivíduos. Os métodos de seleção abordados foram o elitismo e a roleta. O elitismo faz com que os m melhores indivíduos da geração t sejam passados diretamente à próxima, e os indivíduos que faltam (p indivíduos) para completar os n indivíduos da população da geração $t+1$ vêm da seleção da roleta. A roleta seleciona p pares de indivíduos e os submete aos operadores genéticos, que originam os $(n - m)$ novos indivíduos para compor a população. O tamanho da população de cada geração é dada pela equação 2.

$$n = m + p$$
 Equação (2)

Operadores Genéticos: são as formas com que o material genético dos indivíduos são trocados ou modificados no momento da reprodução. Os operadores utilizados foram o cruzamento e a mutação. Baseado no parâmetro “probabilidade de cruzamento”, do algoritmo genético, verifica-se se o mesmo é satisfeito. Caso positivo, ocorre o cruzamento entre os pais selecionados pela roleta e o indivíduo resultante é adicionado à população da próxima geração. Se o parâmetro não for satisfeito, um dos pais selecionados é modificado pelo operador de mutação, criando um novo indivíduo, que por sua vez é adicionado à população da próxima geração.

No cruzamento (figura 2), a composição do indivíduo “filho” é baseada em sorteios (1 ou 0). Se o valor sorteado for 0, a matriz do filho resultante é originária do pai 0, se o valor sorteado for 1, a matriz que fará parte do “filho” é oriunda do pai 1.

Na mutação (figura 3), é verificado cada gene do indivíduo em todas suas matrizes e, conforme o parâmetro “probabilidade de mutação”, os genes podem ou não serem trocados de lugar um com os outros de forma aleatória.



Figura 11. Operadores genéticos: cruzamento entre pais

É importante observar que a distribuição dos genes dentro das matrizes não é alterada pelo operador de cruzamento, e sim pelo operador de mutação. Se não fosse por meio deste operador, a distribuição dos genes dentro das matrizes nunca seria alterada desde a criação da população inicial até o término da execução do algoritmo. O operador de mutação possibilita, desta forma, que a técnica consiga aumentar seu espaço de busca, não previsto em sua primeira geração.

Critério de Parada: Foi utilizado como critério de parada o parâmetro número máximo de gerações ($t_{máx}$). Quando este é alcançado, o que possuir a melhor avaliação e pertencer à população da geração $t_{máx}$ é tido como resposta para o problema.

| | Segunda | Terça | Quarta | Quinta | Sexta |
|-------|---|--|---|--|--|
| Noite | Geometria Analítica Michêl Peters 4109 – Sala de Aula Curso – EC I Sem | Cálculo I Karla Morales 4309 – Sala de Aula Curso – EC I Sem | Algoritmos e Programação Sandra Piovesan 2306 – Lab. Informática Curso – EC I Sem | Introdução à Engenharia de Computação Sandra Piovesan 4211 – Sala de Aula Curso – EC I Sem | Introdução à Arquitetura de Computadores Sandro Camargo 2103 – Sala de Aula Curso – EC I Sem |
| Noite | Produção Acadêmico Científica Cássia 1307 – Sala de Aula Curso – EC I Sem | Introdução à Arquitetura de Computadores Sandro Camargo 2103 – Sala de Aula Curso – EC I Sem | Algoritmos e Programação Sandra Piovesan 4109 – Sala de Aula Curso – EC I Sem | Geometria Analítica Michêl Peters 4109 – Sala de Aula Curso – EC I Sem | Cálculo I Karla Morales 4309 – Sala de Aula Curso – EC I Sem |

| | Segunda | Terça | Quarta | Quinta | Sexta |
|-------|---|--|---|--|--|
| Noite | Algoritmos e Programação Sandra Piovesan 4105 – Sala de Aula Curso – EC I Sem | Cálculo I Karla Morales 4309 – Sala de Aula Curso – EC I Sem | Algoritmos e Programação Sandra Piovesan 2306 – Lab. Informática Curso – EC I Sem | Introdução à Engenharia de Computação Sandra Piovesan 4211 – Sala de Aula Curso – EC I Sem | Introdução à Arquitetura de Computadores Sandro Camargo 2103 – Sala de Aula Curso – EC I Sem |
| Noite | Produção Acadêmico Científica Cássia 1307 – Sala de Aula Curso – EC I Sem | Introdução à Arquitetura de Computadores Sandro Camargo 2103 – Sala de Aula Curso – EC I Sem | Geometria Analítica Michêl Peters 4109 – Sala de Aula Curso – EC I Sem | Geometria Analítica Michêl Peters 4109 – Sala de Aula Curso – EC I Sem | Cálculo I Karla Morales 4309 – Sala de Aula Curso – EC I Sem |

Figura 12. Mutaç o de um indiv duo

3. EXPERIMENTOS E RESULTADOS

Os experimentos foram realizados em uma plataforma *Windows 8.1*, com processador Intel Core I7 e 8 Gb de mem ria RAM, sendo o algoritmo implementado de forma sequencial na linguagem de programa o JAVA.

V rios testes foram realizados na gera o de uma grade hor ria. Foram analisadas as informa es da grade hor ria gerada, das parametriza es utilizadas e dos resultados obtidos pelo algoritmo. Isto possibilitou verificar quais par metros deveriam ser maximizados na fun o de avalia o e os cuidados a serem tomados para que bonifica es n o amortizassem penaliza es. Com isso, chegou-se a configura o de par metros da fun o de avalia o para execu o do algoritmo conforme mostra a Tabela 1. A obten o destes par metros ocorreu de forma emp rica, sendo que a cada execu o do algoritmo, os resultados eram analisados com sua referida parametriza o, e comparados com novas execu es com parametriza es modificadas. Foi realizada a observa o de quanto as modifica es destas vari veis implicariam na avalia o do melhor indiv duo ao final de cada execu o do algoritmo.

Tabela 4. Tabela dos par metros da fun o de avalia o.

| Par metro | Peso atribuido |
|---|----------------|
| Penaliza o por aloca o de docente em per odo indispon vel | 120 |
| Penaliza o sobre o c lculo do desvio padr o na distribui o de disciplinas | 2000 |
| Penaliza o por ocorr ncia de colis o entre salas/hor rios dos professores | 700 |
| B nus por aloca o de docente em per odo preferencial | 30 |

Para simula o, foram importados da planilha de hor rios do semestre 2014/1 do campus Bag  da UNIPAMPA (dispon vel em seu s tio da *internet*) os dados das afinidades entre os docentes e as disciplinas, e as prefer ncias dos dias e per odos que estes possuem para ministrar suas aulas. Isto foi baseado nas aloca es constantes na referida planilha, de forma que se um docente foi alocado, por exemplo, na segunda-feira, nos dois  ltimos per odos da noite e na ter a-feira, nos dois primeiros per odos da tarde, para efeito de simula o, estes s o os seus per odos preferenciais e o mesmo possui afinidade com a referida disciplina. Contabilizou-se um total de 638 per odos preferenciais e foram adicionados aleatoriamente 24 per odos divididos entre 4 docentes, em que os mesmos estariam indispon veis para ministrar suas disciplinas. O

quadro abaixo mostra o resumo de uma das melhores execuções da implementação:

| |
|---|
| Tamanho das populações: 1000 |
| Número máximo de gerações: 3000 |
| Percentual Elitismo: 10 |
| Probabilidade de mutação: 0.5% |
| Probabilidade de Crossover: 65% |
| Avaliação melhor indivíduo: 30709 |
| Total de colisões de salas: 0 |
| Total de colisões de docentes: 0 |
| Total de Disciplinas Alocadas: 248 |
| Total de Períodos Alocados: 429 |
| Desvio padrão na distribuição de disciplinas por docente: 1.91 |
| Total de quebras de restrições de docentes em horários indisponíveis: 0 |
| Total de alocações em horários preferenciais de docentes: 293 |
| Tempo Total: 9402.4 segundos (~ 2 h e 30 min) |

Observa-se que do total de períodos alocados, 68,3% foram em horários que os docentes possuíam preferências por ministrarem suas aulas. Os outros 31,7% eram indiferentes quanto aos períodos destinados a eles. Todos os períodos em que os docentes encontravam-se indisponíveis foram preservados.

Restrições que invalidam a aplicação de uma grade horária, como a alocação de uma sala para duas disciplinas diferentes concomitantemente e um professor ser alocado para ministrar aulas em duas ou mais salas simultaneamente, foram satisfeitas.

Os testes foram realizados gerando as grades horárias para os semestres ímpares de 10 cursos de graduação do campus Bagé da UNIPAMPA, totalizando 248 disciplinas alocadas utilizando para isso 429 períodos disponíveis.

A curva que mostra a evolução das aptidões (avaliações) com a evolução das gerações pode ser observada na figura 4. Nota-se o rápido crescimento da curva, isso acontece por causa das colisões de salas e docentes, que por possuírem um dos maiores valores na parametrização das penalizações, são tratadas primeiramente até serem sanadas.

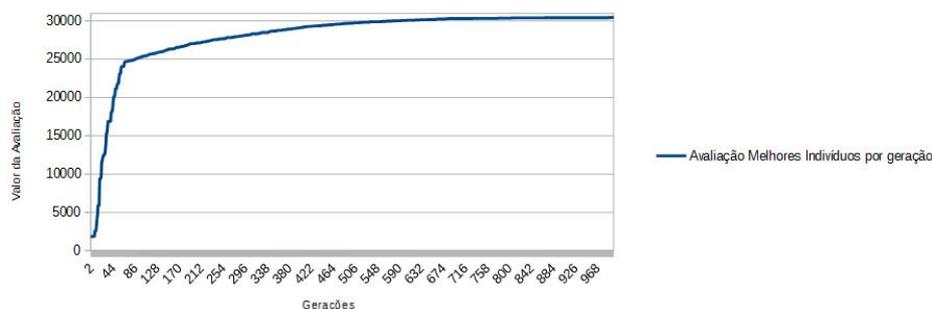


Figura 13. Evolução das avaliações do indivíduos , limitado a mil gerações

Posteriormente pode-se observar um declínio na taxa de crescimento dessa curva, o que nos leva a perceber que o algoritmo faz com que passem a serem exploradas as soluções que, após não possuírem mais colisões entre salas e docentes, coíbem as alocações em que professores estariam indisponíveis para ministrarem suas aulas. Em seguida, busca as soluções que contemplem também as alocações preferenciais dos docentes, que é o parâmetro de bônus com menor peso.

Para confrontar esta informação, pode-se observar o gráfico da figura 5, onde nota-se que tanto a curva das colisões de salas, quanto a curva das colisões de docentes decaem de forma tão rápida quanto a curva da avaliação dos melhores indivíduos cresce na figura 4. Sendo que ainda se observa na figura 5 a queda também acentuada do número de alocações de docentes em períodos em que os mesmos estão indisponíveis.

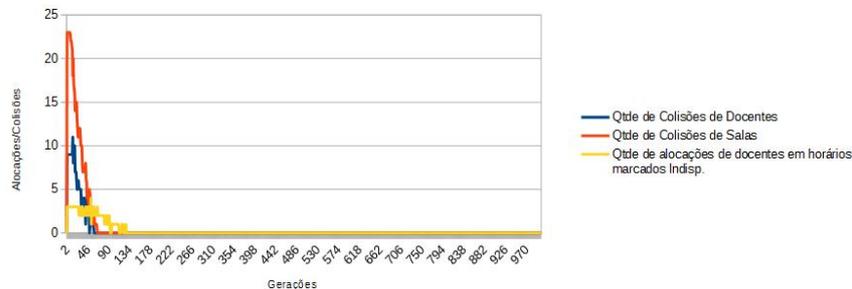


Figura 14. Quantidade de alocações de salas/docentes com colisão, e quantidade de docentes alocados em períodos marcados como indisponíveis,

Quando a curva do melhor indivíduo da figura 4 começa a diminuir sua taxa de crescimento, acontece a busca do AG em alocar os docentes que ainda não estão seus em seus horários preferenciais. Isto pode ser confirmado pela figura 6, que é gráfico que mostra o crescimento do número de alocações dos docentes em horários de sua preferência. Esta figura ainda mostra que no início da execução do algoritmo, as primeiras gerações são passadas sem que o número da quantidade de alocações preferenciais aumente de forma substancial. Este crescimento ocorre logo após este pequeno período e prossegue até o final da execução do algoritmo.

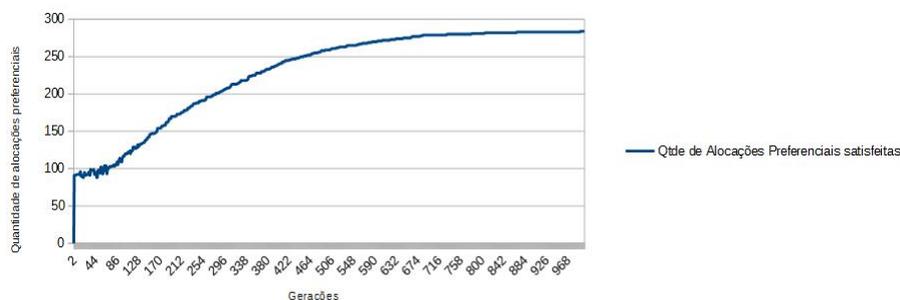


Figura 15. Quantidade de alocações preferenciais satisfeitas, limitado a mil gerações

4. CONCLUSÕES E TRABALHOS FUTUROS

A técnica de exploração de espaço de busca mostrou-se funcional em relação ao problema proposto, sendo que grades horárias aplicáveis foram geradas com sucesso. Grades que não poderiam ser aplicadas (com colisões de salas e horários de docentes) também foram geradas, porém somente ocorreram casos isolados, com populações pequenas, que limitavam muito o espaço de busca do AG.

Pode-se observar que tanto o tamanho da população quanto o número máximo de gerações foram itens impactantes no desempenho do algoritmo, pois o tempo de execução está diretamente relacionado com estes, principalmente por causa do número de vezes com que as funções de cruzamento, mutação e avaliação são executadas.

Todavia, quanto maior o tamanho das populações, maior ser torna o espaço de busca e a diversidade genética. Na maioria dos testes, um espaço de buscas grande

encontrou soluções bem mais adequadas do que com populações pequenas. Algumas simulações com populações de 300 indivíduos e número máximo de 300 gerações apontaram para grades de horários que não possuíam colisões de docentes/salas e nem ocorriam quebras de períodos em que os docentes estariam indisponíveis, porém poucos períodos preferenciais foram alocados. Isto mostra que a partir de um certo número de gerações o algoritmo busca algum tipo de “ajuste fino” para a sua solução, como pôde ser observado nas figuras 4, 5 e 6.

Como trabalho futuro, observa-se a ideia de alocação e divisão dos discentes em turmas, que não foi o foco deste trabalho. Essa divisão, possibilitará a alocação de turmas de diferentes cursos numa mesma sala quando a disciplina entre ambos for igual, e isto ajudaria tanto na alocação de salas, bem como, principalmente na melhor distribuição da carga horária entre os docentes.

Seria interessante a execução desta técnica, de modo a explorar o potencial de arquiteturas paralelas. Como acontece com a função de avaliação, que em uma população de 3000 indivíduos, é executada 3000 vezes de forma sequencial somente para uma única geração, ela poderia ser executada de forma paralela para mais de um indivíduo simultaneamente, obtendo considerável otimização no tempo de execução.

6. REFERÊNCIAS

- Linden, R. (2012) “Algoritmos Genéticos” 3ª Edição. Rio de Janeiro: Editora Ciência Moderna Ltda.
- Gervásio, Jairo de F. (2012) “Computação Evolutiva Aplicada ao Problema de Grade Horária: o Caso do Curso de Análise e Desenvolvimento de Sistemas do IFTM” Dissertação (Mestrado em Ciências), Universidade Federal do Triângulo Mineiro, Uberlândia.
- Lobo, Eduardo Luis Miranda (2005) “Uma Solução Do Problema De Horário Escolar Via Algoritmo Genético Paralelo” Dissertação (Mestrado em Modelagem Matemática e Computacional), Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte.
- Simonetti, Geraldo Bello (2007) “Abordagem do Problema de Programação de Grade Horária Sujeito a Restrições Utilizando Coloração de Grafos” Dissertação (Mestrado em Informática), Universidade Federal do Espírito Santo, Vitória.
- Ferreira, Hélio de L. J.; Corrêa, Marcos V. (2010) “Implementação de uma Sistema de Alocação de Professores e Disciplinas em Grades Horárias Utilizando Algoritmos Genéticos” Trabalho de Conclusão de Curso (Bacharel em Ciências da Computação), Universidade do Anhembi Morumbi, São Paulo.
- De Oliveira, Osmar Braz Júnior (2000) “Otimização de Horários em Instituições de Horários em Instituições de Ensino Superior Através de Algoritmos Genéticos” Dissertação (Mestrado em Engenharia de Produção), UFSC, Florianópolis.