



UNIVERSIDADE FEDERAL DE MINAS GERAIS
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO

GABRIEL MORAES DE SOUZA DINIZ

Avaliação do Perfil Oncológico do Brasil

Proposta de Projeto de Pesquisa Científica
Projeto Orientado em Computação I

Orientador: Adriano Alonso Veloso

SUMÁRIO

1 INTRODUÇÃO.....	3
2 REFERENCIAL TEÓRICO.....	4
3 METODOLOGIA.....	4
4 RESULTADOS ESPERADOS.....	5
5 ETAPAS E CRONOGRAMA.....	6
REFERÊNCIAS BIBLIOGRÁFICAS.....	7

INTRODUÇÃO

Segundo uma previsão realizada pelo International Union Against Cancer (UICC), em 2020 haverá mais de 15 milhões de casos de câncer no mundo. Hoje a doença já é responsável por mais de 12% de todas as causas de óbito no planeta. Muito se investe a fim de se obter uma detecção e prevenção mais efetiva, na descoberta dos principais fatores causadores da doença e nos principais perfis de riscos de cada tipo de câncer.

Diante deste problema, esse projeto científico sugere uma nova análise epidemiológica do câncer e possui a finalidade de obter resultados relevantes acerca do perfil oncológico no Brasil, que será determinado de acordo com o estilo de vida, a idade, os hábitos de saúde e informações genéticas dos casos clínicos por meio dos conceitos computacionais de Aprendizagem de Máquina e processamento de linguagem natural.

Nesta primeira instância do trabalho (POC I), é desejável alcançar uma preparação da abordagem do problema, isto é: realizar inicialmente uma formatação do problema e um devido preparo científico mais preciso para a definição da melhor abordagem computacional de resolução do problema.

Em continuidade a este projeto (POC II), espera-se realizar experimentos com a abordagem definida e realizar várias iterações de melhorias da complexidade e acurácia no algoritmo para que se alcance resultados considerados satisfatórios na pesquisa.

Por fim, ao término do projeto, com os resultados obtidos no agrupamento dos variados perfis para os diversos tipos de doenças, planeja-se extrair conclusões sobre o perfil oncológico no Brasil e comparar com os resultados obtidos em pesquisas já existentes no mesmo campo de pesquisa.

REFERENCIAL TEÓRICO

O risco de ocorrência de uma doença pode ser representado pela probabilidade de uma ocorrência levando em consideração fatores como as condições de vida, sociais, econômicas e saúde de um indivíduo. No caso do câncer, há diariamente novos investimentos em pesquisa para compreender a multiplicidade de causas que determina os riscos de se adquirir tal doença.

Um dos maiores aliados no tratamento do câncer é a detecção precoce, que é o diagnóstico pré-sintomático da doença. Quanto mais cedo a doença é detectada, maiores são as chances de cura e menos agressivo é o tratamento. Segundo a Organização Mundial de Saúde, há duas recomendações para se realizar a detecção precoce: o diagnóstico precoce e o rastreamento de pessoas com maiores riscos. Dessa forma são realizados exames mais específicos com pessoas consideradas suspeitas ou mais propícias a terem a doença de acordo com fatores determinados por dados históricos de pacientes da doença; no caso do Câncer de Mama, por exemplo, mulheres entre 50 e 69 anos são consideradas pacientes de risco e devem realizar exames específicos de rastreamento.

Apesar de existir uma quantidade satisfatória de dados relacionados a pacientes com câncer no país, poderiam existir mais inferências baseadas nestes dados com o propósito de auxiliar na detecção precoce da doença. Assim, uma análise mais minuciosa desses dados é importante para determinação de grupos de pacientes que apresentam maiores riscos da doença baseado em outras semelhanças com os casos existentes.

METODOLOGIA

O primeiro passo para solução desse problema é a aquisição dos dados do Instituto Oncoguia (<http://www.oncoguia.org.br/>), a maior plataforma online que apoia pessoas com câncer e possui uma ferramenta de coleta de dados de casos da doença no Brasil.

A formatação do problema é um dos passos fundamentais desse projeto, posto que a resolução é computacionalmente complexa e exige que haja uma estratégia que precede a execução. Portanto, a escolha da melhor abordagem computacional é um passo importante nesta resolução. Essa estratégia se refere ao estudo dos diversos algoritmos de agrupamento de dados em Aprendizagem de Máquina existentes e a escolha do algoritmo adequado para esse problema específico; além disso, é importante salientar que esse tipo de algoritmo requer uma seleção dos dados que são significantes no contexto do problema e a magnitude da significância desse dado para estabelecer o perfil oncológico.

Após essa etapa de preparação, estudo e análise, inicia-se a etapa executora com a preparação dos dados disponibilizados pelo Oncoguia: estes dados precisam ser filtrados com o que foi julgado importante na etapa anterior e serão divididos em dados de treinamento e dados de teste para o algoritmo. Basicamente os dados de treinamento servirão para que se formule um algoritmo genérico de agrupamento e os dados de testes serão responsáveis para mensurar a qualidade.

Várias iterações de aperfeiçoamento e ajustes serão executadas para que se garanta que a formulação do algoritmo seja independente aos dados treinamentos e que garanta que o algoritmo se comporte de maneira adequada a novos dados desconhecidos.

RESULTADOS ESPERADOS

Espera-se, ao término da primeira parte deste projeto que haja uma formatação completa do problema com o devido preparo científico, a respeito da epidemiologia dos tipos de câncer, e prático, que se refere às práticas computacionais necessárias para o desenvolvimento do projeto na resolução do problema em aprendizagem de máquina.

A primeira etapa do projeto possibilitará também que se estabeleça um perfil demográfico e social de cada tipo de câncer, isto é, que seja realizada uma análise quantitativa do problema que será de suma importância para o restante do projeto porque ela será fundamental para que se possa realizar um entendimento maior dos dados e delimitar os experimentos que serão realizados no POC II.

Ao término do projeto (POC II), após a análise e processamento dos dados por meio de recursos computacionais espera-se obter o perfil adequado para cada tipo de câncer, que deve ser modelado especificamente de acordo as características do tipo da doença. Por exemplo, no caso do câncer de Próstata questões relevantes para modelagem do perfil são: Quais os estados com maiores ocorrências? Quais é a dúvida que geralmente os pacientes têm? Qual a profissão mais comum dos pacientes dessa doença? Qual a faixa de idade dos pacientes? e etc. Portanto, o resultado final do projeto será a modelagem e a determinação de um perfil específico para cada tipo de câncer de acordo com as informações relevantes extraídas do Oncoguia sobre o tipo da doença.

ETAPAS E CRONOGRAMA

SEMANA	TAREFA
01/04/2018 a 07/04/2018	Elaboração da proposta do projeto
08/04/2018 a 14/04/2018	Determinação de características relevantes dos dados
15/04/2018 a 21/04/2018	Estudo de algoritmos de aprendizagem de máquina
22/04/2018 a 28/04/2018	Estudo da epidemiologia dos tipos de câncer dos dados
29/04/2018 a 05/05/2018	Apresentação parcial
06/05/2018 a 12/05/2018	Análise quantitativa dos dados
13/05/2018 a 19/05/2018	Análise quantitativa dos dados
20/05/2018 a 26/05/2018	Estudo de algoritmos de aprendizagem de máquina
27/05/2018 a 02/06/2018	Determinação do algoritmo que será utilizado no projeto
03/06/2018 a 09/06/2018	Gerar representações gráficas dos resultados
10/06/2018 a 16/06/2018	Apresentação Final
17/06/2018 a 23/06/2018	Elaboração do Relatório Final da disciplina
23/06/2018 a 29/06/2018	Relatório Final

REFERÊNCIAS BIBLIOGRÁFICAS

A DETECÇÃO PRECOCE. INCA. Disponível em:

http://www2.inca.gov.br/wps/wcm/connect/acoes_programas/site/home/nobrasil/programa_controle_cancer_mama/deteccao_precoce

Acesso em: 01 abril 2018

CADERNO DE ATENÇÃO PRIMÁRIO. MINISTÉRIO DA SAÚDE. Disponível em:

http://bvsmms.saude.gov.br/bvs/publicacoes/caderno_atencao_primaria_29_rastreamento.pdf

Acesso em: 02 abril 2018

BENGIO, Youshua; GOODFELLOW, Ian; COURVILLE, Aaron; Deep Learning. 1ed: MIT, 2016,

A SITUAÇÃO DO CÂNCER NO BRASIL. INCA. Disponível em:

http://bvsmms.saude.gov.br/bvs/publicacoes/situacao_cancer_brasil.pdf

Acesso em: 02 abril 2018

INSTITUTO ONCOGUIA. Disponível em:

<http://www.oncoquia.org.br/>

Acesso em: 02 abril 2018