

Universidade Federal de Uberlândia
Faculdade de Computação

VII FACOM TECHWEEK E XIV WORKSHOP DE TESES E DISSERTAÇÕES EM CIÊNCIA DA COMPUTAÇÃO

Anais

09 a 20 de novembro de 2020

ISSN: 2447-0406



FACOM
TECHWEEK

WTDC
XIV Workshop de Teses e Dissertações
em Ciência da Computação

Uberlândia
2020

VII FACOM Techweek

Anais do XIV WTDC

Evento virtual

09 à 20 de novembro de 2020

Comissão Organizadora

Carlos Ramon Pantaleon Dionísio (UFU)

Elaine Ribeiro de Faria Paiva (UFU)

João Batista Simão (UFU)

Murillo Guimarães Carneiro (UFU)

Renan Gonçalves Cattelan (UFU)

Renato de Aquino Lopes (UFU)

Wendel Alexandre Xavier de Melo (UFU)

Organização, Execução, Promoção e Realização

Universidade Federal de Uberlândia (UFU)

Faculdade de Computação (FACOM/UFU)

FACOM Techweek e WTDCC

A TechWeek é um evento realizado pela Faculdade de Computação (FACOM) da Universidade Federal de Uberlândia (UFU), que busca promover atualização técnica e apresentar as últimas tendências tecnológicas, reunindo e proporcionando uma maior interação entre estudantes, professores, profissionais e empresas das áreas de Tecnologia da Informação. Trata-se de um evento tradicional, já em sua sétima edição, composto por palestras, minicursos, mesas redondas, competições técnicas e apresentação de trabalhos e pesquisas científicas.

Realizado em conjunto com a Techweek desde 2016, o Workshop de Teses e Dissertações em Ciência da Computação (WTDCC) é um evento do Programa de Pós-Graduação em Computação (PPGCO) da UFU, na sua décima quarta edição, o qual conta com a participação de estudantes de graduação e pós-graduação de universidades de Uberlândia e região. Dentre os objetivos do WTDCC destacam-se: incentivar a troca de experiência entre pesquisadores, apoiar os trabalhos produzidos pelos discentes, promover a interação entre os diversos grupos de pesquisa e refletir sobre a atual situação deles no cenário nacional. Nesse sentido, o evento consiste em um importante espaço de integração para a troca de experiências acadêmico-científicas, objetivando o desenvolvimento da ciência e da tecnologia nesse domínio do conhecimento.

Considerando as recomendações de isolamento social em virtude da pandemia do COVID-19, ambos os eventos aconteceram pela primeira vez de maneira online.

Prefácio

O Workshop de Teses e Dissertações em Ciência da Computação foi realizado junto com a VII FACOM Techweek e teve como objetivo divulgar os projetos de pesquisa científica e tecnológica na área de computação, realizados por estudantes de graduação e pós-graduação da Universidade Federal de Uberlândia (UFU) e região. Também teve por objetivo contribuir para a formação dos participantes despertando o interesse pelas descobertas científicas e pela resolução de problemas complexos.

Na sua décima quarta edição, o WTDCC contou com palestras temáticas, bem como a submissão e apresentação de trabalhos. As palestras compreenderam tópicos relacionados às linhas de pesquisa desenvolvidas pelo Programa de Pós-Graduação em Ciência da Computação (PPGCO/UFU), bem como sobre os desafios da pós-graduação e a situação atual do programa em nível nacional e internacional. As submissões de trabalho, em formato de resumo, contemplaram três trilhas relacionadas à formação dos discentes: trabalhos de graduação, mestrado e doutorado, vários deles apresentados durante o evento em sessões técnicas virtuais.

Além de contar com os trabalhos de dezenas de autores, o WTDCC também contou com boa presença do público, interessado em conhecer mais sobre os trabalhos desenvolvidos em Uberlândia e região. Nesse sentido, agradecemos a todos os alunos e orientadores que submeteram seus trabalhos para o evento e também ao público que prestigiou as palestras e apresentação de trabalhos.

Por fim, agradecemos à comissão de avaliação dos trabalhos e também às empresas e instituições de Uberlândia e região que marcaram presença no evento.

Profa. Elaine Ribeiro de Faria Paiva

Prof. Murillo Guimarães Carneiro

Coordenadores do XIV WTDCC

Organização

Coordenação Geral

Prof. Renan Gonçalves Cattelan (UFU)

Coordenação do XIV WTDC

Profa. Elaine Ribeiro de Faria Paiva (UFU)

Prof. Murillo Guimarães Carneiro (UFU)

Comissão de Avaliação dos Trabalhos

Prof. Carlos Ramon Pantaleon Dionísio (UFU)

Prof. Renato de Aquino Lopes (UFU)

Equipe de Trabalho

Anagê Calixto Mundim Filho

Bruna Alves Oliveira

Bruno de Paula Nery

Dahlan Pereira Gardim

Diogo Honório Rodrigues Silva

Felipe Augusto Ferreira de Castro

Gabriel de Araújo Velasco

Gabriel Teodoro Ribeiro

Gabriel Vinicius Ramos de Lacerda

Giullia Rodrigues de Menezes

Higor Raphael Faria e Sousa

Kelly Cristina Alves

Laís Saloum Deghaide

Lucas Fernando Borges

Lucas Guimarães Mendes

Marcelo Mendonça Borges

Mateus de Freitas Rosa

Nayara Terezinha Nunes

Roberto Costa Tupinambá

Rodrigo Zamboni Silva

Salomão Oliveira Alves

Yan Gustavo Pegyn Silva

Yan Stivaletti E Souza

Editoração dos Anais

Profa. Elaine Ribeiro de Faria Paiva (UFU)

Prof. Murillo Guimarães Carneiro (UFU)

Contents

I TRABALHOS DE GRADUAÇÃO	1
Redes convolucionais para classificação e avaliação da maturação do cafeeiro	2
Extração de Características Aprendidas para Classificação de Lesões da Cavidade Bucal	3
Avaliação comparativa entre medidas de redes complexas para a classificação de dados	4
Aplicação de técnicas de agrupamento em dados de acidentes de trabalho	5
Simulando urnas de Pólya	6
Desenvolvimento de Plataforma IoT com comunicação LoRa para Auxiliar no Planejamento do Serviço de Coleta de Lixo	7
Inclusão Divertida - Idosos Aprendendo Informática com Jogos	8
Experimentações em Sistemas Generativos para Produção de Terrenos	9
Mineração de texto aplicada a relatos de violência contra a mulher	10
Investigação de fraudes financeiras eletrônicas usando mineração de dados	11
Desenvolvimento de uma Plataforma Web Didática para auxiliar o Ensino de Estrutura de Dados	12
Análise de métodos de construção do grafo para detecção de comunidades	13
Análise de vulnerabilidades de segurança usando técnicas de mineração de dados . .	14
Estudo da linguagem Rust na construção de aplicações embarcadas	15
Redes Neurais Artificiais para modelagem e previsão de dados meteorológicos da região do Alto Paranaíba do Triângulo Mineiro	16
II TRABALHOS DE MESTRADO	17
Imagens infravermelhas e redes convolucionais para detecção de câncer de mama . .	18
Uma proposta para implementação prática de auto-configuração especializada de elementos de borda em redes SDN Utilizando a arquitetura SONAR	19
Dispositivo de monitoramento e acompanhamento de frequência cardíaca e temperatura corporal	20
Detecção de Intrusão sobre Pacotes de Redes utilizando Algoritmos de Fluxos Contínuos de Dados	21
Nova heurística de fluxo de informação para classificação de alto nível usando redes complexas	22
Algoritmos evolutivos multiobjetivo e modelo ETC para escalonamento de tarefas em ambientes distribuídos	23
Detecção e correção de distorções não lineares para co-registro multiespectral de imagens agrícolas obtidas por veículos aéreos não tripulados	24
Aprendizado de padrões em sinais EEG para apoio ao prognóstico de pacientes em coma	25
Processamento e Classificação de Séries Temporais Aplicados ao Desafio PLAsTiCC .	26
Um Novo Algoritmo De Difusão De Distância Para Um Modelo De Planejamento De Caminho Baseado Em Autômatos Celulares	27
Comparação de descritores morfológicos e não morfológicos na classificação multi classe de imagens histológicas de linfomas não Hodgkins	28
III TRABALHOS DE DOUTORADO	29
Michelzinho: Jogo sério para o ensino de habilidades emocionais em pessoas com autismo ou deficiência intelectual	30
Classificação de Lesões Pré Cancerosas em Imagens Histológicas da Cavidade Oral . .	31
Mineração de Soluções Relevantes para Tarefas de Programação a partir de Resultados de Serviços de Busca	32

Automated Extraction of Tumor Region of Interests from H&E-stained Oral Histological Whole Slide Images	33
Efficient Data Structures for Caching and Querying Large Temporal Graphs	34

Chapter I

TRABALHOS DE GRADUAÇÃO

Redes convolucionais para classificação e avaliação da maturação do cafeeiro

Anagê C. Mundim Filho (Universidade Federal de Uberlândia); Cleyton B. Alvarenga (Universidade Federal de Uberlândia); Murillo Guimarães Carneiro (Universidade Federal de Uberlândia)

anagefilho@gmail.com; cleytonalvarenga@ufu.br; mgcarneiro@ufu.br

Resumo: Este projeto tem o intuito de desenvolver um sistema computacional para auxiliar os produtores de café na escolha do melhor momento para colheita dos grãos. Apesar de ser fonte de grande riqueza para nosso país, o café ainda é colhido com bastante incerteza por conta da variação no teor de água dos frutos. A metodologia mais comum para a colheita costuma ser a tomada de decisão por parte do agricultor após uma análise visual do talhão, seguida pela observação dos frutos ou do desprendimento dos mesmos. Como forma de automatizar o processo e garantir uma melhor tomada de decisão por parte do produtor, o presente projeto propõe investigar técnicas computacionais de Inteligência Artificial denominadas Redes Neurais Convolucionais, as quais serão treinadas com milhares de fotos de talhas de café em seus cinco estádios de maturação: verde; verde cana; cereja; passa; seco. Diferente de trabalhos relacionados que buscam apenas classificar os frutos, este projeto visa o desenvolvimento de modelos capazes de também auxiliar o produtor quanto ao momento mais adequado para a colheita. O modelo treinado deverá, a partir de imagens das plantas de café capturadas pelos próprios produtores, retornar se estes devem ou não realizar a colheita de tais frutos, auxiliando-os na tomada de decisão. Especificamente, almeja-se comparar o desempenho preditivo da abordagem desenvolvida com duas redes convolucionais bastante conhecidas no contexto de classificação das imagens: LeNet-5 e VGG-16. Para auxiliar no treinamento dos modelos, foi desenvolvida uma metodologia para seleção e anotação das imagens de café, a qual busca caracterizar os estádios de maturação dos frutos através de diversas anotadas por vários especialistas da área. Apesar do trabalho ainda estar em andamento, alguns resultados preliminares indicam que a abordagem é promissora e que, a partir de um aplicativo para dispositivo móvel, poderá contribuir para uma colheita de maior qualidade por parte de pequenos e médios produtores

Trilha: Trabalhos de Graduação

Extração de Características Aprendidas para Classificação de Lesões da Cavidade Bucal

André D. Freitas (Universidade Federal de Uberlândia); Marcelo Zanchetta do Nascimento (Universidade Federal de Uberlândia); Adriano B. Silva (Universidade Federal de Uberlândia); Paulo Rogério de Faria (Universidade Federal de Uberlândia)

andredf@ufu.br;

marcelo.zanchetta@gmail.com;

adrianobs@gmail.com;

paulorfaria1976@gmail.com

Resumo: O carcinoma de células escamosas bucal é o sétimo tumor maligno mais comum, com mais de 15.000 novos casos por ano no Brasil. Em histologia, os principais métodos para diagnosticar a doença consistem em analisar amostras de tecido coradas com hematoxilina e eosina, o que pode ser impreciso para a classificação. Sistemas de apoio ao diagnóstico (do inglês, computer aided diagnosis – CAD) são empregados para auxiliar o especialista na tomada de decisão durante o diagnóstico para a classificação de anormalidades. Nesse sistema, os descritores de características cuja função é representar informações dos pixels em um conjunto de dados. A etapa de extração de característica é importante para o sucesso da execução dos algoritmos de classificação. Neste trabalho é apresentado uma aproximação para a análise de imagens histológicas da cavidade bucal com descritores aprendidos e classificação de diferentes classes de lesões da cavidade bucal (Saudável, Leve, Moderado e Severo). Nesse estudo, foram usados os modelos de redes neurais AlexNet, VGG16 e ResNet para extração das características na camada de Flatten. A partir desses modelos, foram extraídas e selecionadas as características aprendidas de cada imagem e foram avaliadas com os algoritmos de aprendizado de máquina Random Forest (RF) e SVM (Support Vector Machine). As métricas Sensibilidade, Especificidade, Precisão e Acurácia foram empregadas para avaliação dos modelos. Com uso, do modelo ResNet e classificador SVM o valor médio de acurácia para os grupos foram de 98,4%. No caso da classe saudável notou-se que a acurácia foi de 99,4%. Já no caso de uma classificação apenas usando os modelos de redes neurais convolucionais o melhor resultado médio foi a ResNet com 95,9%. Nessa avaliação, o grupo de lesões severas foram as que obtiveram o melhor desempenho na classificação (acurácia de 97,4%). Nota-se que o uso do modelo de redes neurais convolucionais ResNet junto com o algoritmo de aprendizado de máquina SVM obtiveram melhores desempenho na comparação investigadas. O uso de modelos de redes em conjunto com algoritmos de aprendizado de máquina contribuiu para investigação de lesões de tecidos da cavidade bucal e podem ser empregados com parte de um sistema CAD.

Trilha: Trabalhos de Graduação

Avaliação comparativa entre medidas de redes complexas para a classificação de dados

Barbara Gama (Universidade Federal de Uberlândia); Murillo Guimarães Carneiro (Universidade Federal de Uberlândia)

barbaracgama@ufu.br; mgcarneiro@ufu.br

Resumo: Redes complexas são redes com estruturas de conexão não triviais, as quais não seguem um padrão regular nem aleatório. O uso delas para classificação de dados tem se tornado tópico de pesquisa cada vez mais relevante na literatura, com vários trabalhos explorando de maneira bem sucedida medidas e propriedades de redes complexas. Na classificação de dados, o uso de redes complexas possibilita interpretar e modelar as relações semânticas entre os dados, permitindo a detecção de padrões estruturais e topológicos relacionados, por exemplo, à formação de padrão dos dados. Várias medidas de redes complexas já foram utilizadas nesse sentido, porém, falta na literatura um estudo sistemático capaz de caracterizar a influência e o comportamento de tais medidas no contexto da classificação de dados. Como existem muitas medidas de redes complexas que possuem características diferentes, este trabalho consiste em desenvolver um ambiente experimental a fim de analisar e comparar o comportamento de algumas medidas, revelando cenários mais adequados para usar uma ou outra medida. Especificamente, o ambiente desenvolvido contempla a construção da rede a partir da rede k-vizinhos mais próximos (kNN) e kNN-seletivo, e a exploração de medidas de redes complexas a partir da técnica de classificação baseada em conformidade de padrão (classificação de alto-nível). Para isso, foram selecionadas oito medidas de redes complexas, a saber: assortatividade, coeficiente de agrupamento local médio, grau médio, intermedialidade, menor caminho médio, proximidade, coeficiente global de agrupamento e centralidade de autovetor. Para as análises, foram consideradas tanto bases artificiais quanto reais, todas na forma de vetor de atributos. Os resultados obtidos apontam que medidas, tais como, menor caminho médio, além de apresentarem melhor desempenho preditivo nos cenários com maior nível de ruído, também são mais robustas à variação do parâmetro de construção da rede. Em suma, esta pesquisa consiste em uma importante iniciativa para auxiliar outros trabalhos relacionados na seleção das medidas de redes complexas para classificação de dados.

Trilha: Trabalhos de Graduação

Aplicação de técnicas de agrupamento em dados de acidentes de trabalho

Bruna A. Oliveira (Universidade Federal de Uberlândia); Daniela F. Giacomelli (Universidade Federal de Uberlândia); Elaine Ribeiro Faria (Universidade Federal de Uberlândia)

brunaolialwes@ufu.br; danielafgiacomelli@gmail.com; elainerfp@gmail.com

Resumo: Com o advento da Internet, muitos dados públicos, antes guardados em grandes arquivos físicos, começaram a ser salvos na rede e disponibilizados para a população. Um desses dados são os Comunicados de Acidente de Trabalho (CAT), que estão disponíveis no Observatório de Segurança e Saúde no Trabalho. Considerando apenas 2018, foram registrados mais de 628 mil CAT's. Cada um desses registros é composto por um conjunto de atributos, dentre os quais pode-se destacar: idade, sexo, local do acidente, tipo de acidente, parte do corpo atingida, data, região do país e classe de atividade econômica. Como a base de CAT's é muito grande, e contém diferentes acidentes de trabalho de diferentes regiões do país, torna-se inviável a sua análise manual. A busca por grupos de acidentes de trabalho com características semelhantes é uma importante tarefa a ser realizada nessa base, visto que ela pode auxiliar na caracterização dos acidentes e conseqüentemente na adequação das políticas públicas para o problema. Este trabalho propõe aplicar e avaliar uma técnica de aprendizado não-supervisionado em dados sobre acidentes de trabalho, a fim de buscar padrões que caracterizem esses acidentes. O algoritmo usado neste trabalho é o COBWEB, já que o mesmo consegue lidar com dados categóricos e numéricos, os quais estão presentes na base de dados de CAT's. O COBWEB é um algoritmo hierárquico probabilístico que é capaz de lidar com grandes quantidades de dados, fazendo apenas uma varredura na base de dados. A qualidade do agrupamento foi avaliado usando a medida de validação do coeficiente da silhueta. O COBWEB é um algoritmo hierárquico que gera como resultado uma árvore de grupos. Como nessa árvore perto da raiz estão os grupos mais gerais, e perto das folhas os mais específicos, foi necessário usar um método para decidir onde cortar a árvore. Se apenas as folhas fossem consideradas, teríamos um grupo muito específico formado apenas por uma instância, ou um grupo de instâncias iguais. Neste trabalho, também foi necessário usar também um algoritmo para verificar se o COBWEB é sensível a ordem de entrada dos dados. Para isso foi utilizado a medida HAI (Hierarchy Agreement Index) que compara duas hierarquias e resulta em um valor que indica o quanto elas são semelhantes. Os resultados indicam que as diferentes hierarquias produzidas pelo COBWEB com o mesmo conjunto de dados, apresentando em ordem diferente, são semelhantes. Logo o COBWEB não se mostrou sensível à ordem de entrada dos dados. Os resultados do agrupamento identificou que homens que trabalham com máquinas e equipamentos, com idade média de 33 anos, são suscetíveis a se machucar trabalhando na pesca e na aquicultura, no Brasil. Este trabalho foi realizado em cooperação com a aluna de mestrado Daniela Freitas Giacomelli.

Trilha: Trabalhos de Graduação

Simulando urnas de Pólya

Felipe Augusto Ferreira de Castro (Universidade Federal de Uberlândia); Rodrigo Lambert (Universidade Federal de Uberlândia)

felipe.castro@ufu.br; rodrigolambert@ufu.br

Resumo: Estudar a evolução de modelos dinâmicos é um problema clássico na ciência. Um caso particular bem conhecido na literatura de probabilidade e estatística é o modelo de Urnas de Pólya. Sua dinâmica consiste em adicionar bolas coloridas a uma urna, seguindo regras determinadas pela assim chamada matriz de reposição, e observar a configuração de cores na urna ao longo do tempo. Por sua versatilidade, tal modelo constituiu uma poderosa ferramenta que possui diversas aplicações práticas. Um processo estocástico é uma sequência de variáveis aleatórias indexadas por um conjunto de índices, que representa, por exemplo, a evolução temporal de um sistema. Exemplos de importância prática são: a flutuação da corrente em um circuito elétrico na presença do assim chamado ruído térmico; as mudanças aleatórias no nível de sinais de rádio sintonizados na presença de distúrbios meteorológicos; e o fluxo turbulento de um líquido ou gás. A estes pode-se adicionar muitos processos industriais acompanhados por flutuações aleatórias e também certos processos encontrados em geofísica (por exemplo, variações no campo magnético da Terra), biofísica (por exemplo, variações nos potenciais elétricos do cérebro registrados em um eletroencefalograma), economia, e Teoria da Informação (por exemplo, a codificação de sequências que inclui a compressão de dados). A área de probabilidade e processos estocásticos constitui, portanto, uma ferramenta essencial para estudo de modelos que possuam algum tipo de comportamento aleatório. Nesse contexto, o modelo de urnas se torna uma notável porta de entrada para o universo dos modelos probabilísticos discretos. Entender um modelo de urnas como um processo estocástico a tempo discreto, e a convergência das proporções de cores de bolas como o estudo de seus teoremas de convergência constituem uma imersão bastante intuitiva no mundo da probabilidade. O objetivo do projeto é desenvolver algoritmos para simulação de urnas de pólya. Tais algoritmos buscam generalidade com o objetivo de atender a maior quantidade de modelos de urnas possível. Como exemplo, as urnas de Eggenberger, Friedman e Bagchi-Pal. Além disso, também serão desenvolvidos algoritmos para simular urnas que possuam a propriedade de lapso de memória.

Trilha: Trabalhos de Graduação

Desenvolvimento de Plataforma IoT com comunicação LoRa para Auxiliar no Planejamento do Serviço de Coleta de Lixo

Gabriel A. Velasco (Universidade Federal de Uberlândia); Rafael Pasquini (Universidade Federal de Uberlândia)

gabrielvelascocam@ufu.br; rafael.pasquini@ufu.br

Resumo: Um dos principais desafios atuais em termos da administração das cidades está relacionado com o planejamento adequado dos serviços essenciais. De maneira geral, as cidades estão em constante expansão, fator este que impõe desafios ao planejamento e gerenciamento efetivos destes serviços. Problemas de planejamento e gerenciamento se manifestam nas mais variadas formas, desde o desperdício de recursos financeiros, até a precarização/degradação dos serviços essenciais. Dentre uma gama de serviços, todos extremamente importantes, este trabalho visa desenvolver um piloto do serviço de coleta de lixo com as tecnologias IoT (Internet of Things - Internet das Coisas, em português) e LoRa (Long Range - Longo Alcance, em português). A rede de comunicação LoRa foi desenvolvida pela Semtech, empresa situada na Califórnia, com o objetivo de suportar comunicação por longas distâncias (cerca de 10km em áreas rurais) visando um consumo baixo de energia. A tecnologia ainda está em desenvolvimento e em 2018 foi lançado um novo módulo que operava em uma área de transmissão maior, com um consumo de energia menor do que o da geração anterior, o que indica que a cada geração a tecnologia tende a ficar mais efetiva. A razão para a escolha IoT e LoRa na coleta de lixo justifica-se pela possibilidade dos pontos de coleta se modificarem com certa dinamicidade, as lixeiras podem ser deslocadas, bem como os atores envolvidos, especialmente a equipe de coleta que viaja a cidade em caminhões. Esperamos que o piloto a ser desenvolvido neste trabalho possa melhorar o funcionamento geral do serviço, coletando lixo de maneira mais efetiva e resultando em um uso mais adequado dos recursos. O cenário contempla lixeiras com sensores capazes de avaliar a taxa efetiva de ocupação das lixeiras, bem como coletar a localização destas, transmitindo estas informações a um sistema de planejamento que, por sua vez, utilizará estas informações para determinar as rotas mais eficazes para que os caminhões façam a coleta do lixo. Os desafios incluem construir o sistema de sensoriamento, transmitir os dados na rede LoRa e implementar a plataforma central para determinação das rotas. Esperamos que a tecnologia possa ser demonstrada na prática através do piloto e que os resultados contribuam, dentre outras coisas, com a melhoria da qualidade de vida nas cidades.

Trilha: Trabalhos de Graduação

Inclusão Divertida - Idosos Aprendendo Informática com Jogos

Giullia R. Menezes (Universidade Federal de Uberlândia); Renan G. Cattelan (Universidade Federal de Uberlândia)

giulliar@gmail.com; renan@ufu.br

Resumo: A exclusão digital é um problema que aflige parte significativa da população, com incidência maior entre o público idoso, justamente um dos segmentos que mais cresce considerando-se o aumento da expectativa de vida. Com o objetivo de auxiliar o público idoso a se adaptar às novas tecnologias e poder delas usufruir, o CompPET desenvolve diversas iniciativas para promover sua inclusão digital. Em geral, a metodologia consiste na oferta de oficinas voltadas ao público idoso, com as seguintes características: flexibilidade parcial de conteúdo; material didático próprio, gratuito e livremente disponível, complementado por conteúdo criado por especialistas do NIC.br e voltado ao público idoso; aulas dinâmicas, com colaboração em pares; aproveitamento de infraestrutura dos laboratórios da universidade; emprego de recursos de acessibilidade; e formação de parcerias. Este projeto tem procurado renovar tal metodologia, buscando melhorar o aprendizado e permitindo que o conteúdo repassado seja realmente absorvido e que os idosos se capacitem para utilizar a tecnologia em seu cotidiano. Foi com a tentativa de renovar esse modelo de ensino, que se notou um grande interesse em comum aos participantes da oficina: os jogos como forma de aprendizagem. A ideia de aprender brincando deu uma nova cara ao projeto e tem moldado a forma de apresentação de seu conteúdo. A autora já participou das aulas oferecidas na oficina como monitora e pôde perceber que o maior desafio do idoso não é lembrar o que se foi passado, uma vez que seu costume de anotar as informações passadas ajuda muito na memorização e consulta, quando necessário. Um grande desafio durante as aulas tem sido o uso do mouse, por se tratar justamente de um dispositivo de manipulação indireta, onde a ideia de um meio para se movimentar dentro de uma tela a partir de movimentos realizados em outro plano (mesa) causa dificuldades para os idosos. Funções como desligar o computador de forma correta, que envolve mover o mouse até o botão Iniciar, clicar e depois mover novamente até a opção Desligar, gerando um novo clique, também se tornam difíceis devido aos diversos movimentos e coordenação motora necessários. Segundo relatos, foi possível perceber que, durante aulas com software de desenho como o Paint, tais tarefas se tornavam menos tediosas e estressantes. Daí surgiu a ideia de incluir jogos nas aulas para que as mesmas se tornassem mais lúdicas e motivantes. Foi feito um teste ad hoc passando-se um jogo em forma de labirinto para que os idosos treinassem o movimento e o controle associados ao uso do mouse, com grande aceitação entre os idosos. O novo desafio do projeto é o estudo de jogos adequados para trabalhar não somente a coordenação motora, mas também raciocínio lógico e memorização.

Trilha: Trabalhos de Graduação

Experimentações em Sistemas Generativos para Produção de Terrenos

Higor Raphael Faria e Sousa (Universidade Federal de Uberlândia); Daniel Abdala (Universidade Federal de Uberlândia)

higorrfsousa@gmail.com; abdala@ufu.br

Resumo: Como consequência do aumento sistemático da produção de jogos digitais e de animações 3D, houve um incremento da necessidade de criação de mundos virtuais ricos em detalhes e coesos, uma atividade que demanda profissionais qualificados, várias milhares de horas de trabalho e, conseqüentemente, muito dinheiro. Neste contexto surge uma demanda mercadológica para encontrar meios que sejam mais rápidos e mais baratos para a produção desses mundos virtuais e que ao mesmo tempo não diminua o controle dos artistas sobre a obra criada. Uma solução que vem ganhando destaque são os sistemas generativos procedurais. Sistemas generativos procedurais são sistemas algorítmicos que criam automaticamente conteúdos específicos a partir de modelos e parâmetros. Como exemplo, cabe citar jogos como Minecraft e No Man's Sky, que empregam geração procedural de conteúdo na geração de seus mundos virtuais. Estes sistemas possibilitam uma grande redução na crescente carga de trabalho de artistas e, conseqüentemente, uma queda no preço de produção e de compra de ativos digitais (como imagens e objetos 3d), aliviando o trabalho de modelagem e tornando mais barata e simples a criação de jogos digitais e de animações. Mesmo com todas as vantagens que o uso de sistemas generativos procedurais traz, tais métodos não são empregados ainda em larga escala. Há três razões principais para isso acontecer. São elas: a) métodos tradicionais de geração procedural de conteúdo são frequentemente complexos e não intuitivos para uso; b) métodos tradicionais de geração procedural de conteúdo são difíceis de controlar; e c) os resultados gerados por métodos procedurais diferentes não são integrados facilmente em um mundo virtual completo e consistente. É necessário portanto fornecer uma ferramenta que faça uso da geração procedural de conteúdo, mas que permita ao usuário se concentrar mais no que deseja criar e não em como isso será criado. Esse trabalho pretende estudar a integração entre métodos procedurais e edição manual de mundos virtuais para o desenvolvimento dessa ferramenta. O objetivo primeiro é a geração de um sistema híbrido (procedural com ajustes manuais) para a geração procedural de terrenos a serem utilizados como assets na composição de mundos virtuais. Ao final do processo de pesquisa e de desenvolvimento, é esperado que se chegue a um software que auxilie no processo de geração de modelos 3D de terrenos em duas etapas principais, sendo a primeira procedural e obrigatória; e a segunda manual e opcional. Na primeira etapa, a partir de dados fornecidos pelo usuário, será gerado um terreno proceduralmente e, então, o usuário poderá optar por alterar os dados de entrada e repetir o processo, caso o resultado não seja satisfatório, ou por seguir para a etapa de edição manual. Na segunda etapa, o usuário pode aplicar ou não edições manuais ao modelo 3D do terreno como preferir até chegar ao resultado que julgar adequado.

Trilha: Trabalhos de Graduação

Mineração de texto aplicada a relatos de violência contra a mulher

Isabella Tannús Corrêa (Universidade Federal de Uberlândia); Elaine Ribeiro de Faria Paiva (Universidade Federal de Uberlândia)

tannusisa@gmail.com; elaine@ufu.br

Resumo: O índice de violência no Brasil é muito alto, em especial quando se considera a violência contra mulheres. A cada 2 segundos, uma mulher é vítima de violência física ou verbal no Brasil, e a cada 1,4 segundo uma mulher é vítima de assédio. A violência contra as mulheres começou a ser vista como um problema público com o crescimento da antropologia e noções do feminismo. Visto que muitos casos de violência contra mulheres se tornam públicos, e tomam muita relevância nas mídias sociais, uma importante questão a ser respondida é se existe algum padrão nesses relatos. A busca por padrões nesses dados pode ser feita usando Mineração de Texto, uma extensão da Mineração de Dados, que é um processo de encontrar correlações e padrões em grandes conjuntos de dados. Este trabalho propõe analisar dados de relatos de violência contra a mulher em busca de padrões que possam caracterizar a agressão ou o agressor. Para isso, dados foram coletados de jornais eletrônicos, sites e redes sociais contendo relatos de vítimas do sexo feminino, independente do tipo de violência sofrida: física, psicológica, sexual, patrimonial ou moral. O pré-processamento dos dados envolveu os seguintes passos: remoção de stopwords; remoção de caracteres não alfabéticos, pontuação, acentuação; conversão de letras maiúsculas para minúsculas; substituições de gírias e abreviações; e lematização. Nuvens de palavras foram usadas para auxiliar na validação do pré-processamento. Para a busca por padrões nos relatos das vítimas foram usadas nuvens de palavras e o modelo estatístico para modelagem de tópicos Latent Dirichlet Allocation (LDA). Com o objetivo de dividir a base de dados em duas classes, agressões constantes e esporádicas, e assim facilitar a busca por padrões, o algoritmo de classificação Naive Bayes foi usado. A partir da aplicação do método proposto aos relatos encontrados na internet, algumas conclusões foram obtidas. O Twitter não foi uma boa fonte de informação porque grande parte dos tweets fugiam do propósito da pesquisa, constituindo-se de brincadeiras e comentários sobre o assunto. Já os textos de jornais eletrônicos se mostraram uma boa fonte de informação. Foi observado que a maior parte dos depoimentos recolhidos são associados a alguma agressão realizada por alguém próximo da vítima, como familiar, cônjuge ou amigo. Outro ponto importante a destacar é que esses casos de violência são, muitas vezes, constantes. O modelo gerado pelo algoritmo Naive Bayes alcançou 71,2% de acurácia na classificação dos relatos como constante ou esporádico. Trabalhos futuros incluem uma parceria com alguma ONG de combate a violência contra a mulher a fim de criar uma nova base de dados para aplicação do método proposto. Também pretende-se refinar o método proposto em busca de novos padrões que possam caracterizar os casos de violência contra a mulher.

Trilha: Trabalhos de Graduação

Investigação de fraudes financeiras eletrônicas usando mineração de dados

Laís S. Deghaide (Universidade Federal de Uberlândia); Rodrigo Miani (Universidade Federal de Uberlândia)

laisdeggaide@gmail.com; rodrigomiani@gmail.com

Resumo: O desenvolvimento da Internet tem facilitado o acesso às informações. Esse fator foi decisivo para o aumento significativo atual das transações financeiras realizadas em todo o mundo. Assim, surgiu o comércio eletrônico e com ele as transações eletrônicas. Este tipo de mercado facilitou e acelerou o processo de transações cotidianas, oferecendo comodidade e preços mais atrativos, sendo viável realizar compra e venda online. Porém, com a expansão deste tipo de comércio novos problemas surgiram, como as fraudes eletrônicas financeiras. Ações eletrônicas fraudulentas são recorrentes ao redor de todo mundo, sendo um problema significativo e de custo considerável para bancos e emissoras de cartões de crédito. Diante desse cenário, a detecção de fraudes tornou-se uma atividade essencial para diminuir o impacto das transações fraudulentas. Baseando-se neste problema comum, o presente projeto consistiu em analisar um conjunto de dados de transações financeiras, classificando-as em normais ou fraudulentas, tendo como objetivo geral utilizar técnicas de mineração de dados para criar modelos de classificação para transações fraudulentas. Para o estudo desse problema, foi necessário aprofundar os conhecimentos sobre mineração de dados, identificar os principais trabalhos da literatura correlata e utilizar ferramentas computacionais que implementam processos de mineração de dados. Essa pesquisa utilizou o conjunto de dados “Synthetic Financial Datasets For Fraud Detection” com o método de aprendizado supervisionado e classificação binária. Também foi dado um foco na utilização de algoritmos de classificação do tipo ensemble. Com base nos resultados de pesquisas anteriores, esses algoritmos possuem um resultado melhor no âmbito de detecção de fraudes financeiras eletrônicas. Sendo assim, este estudo realizou uma análise comparativa com os resultados obtidos por algoritmos tradicionais (Regressão Logística e Naive Bayes) e ensemble (LightGBM, XGBoost e Floresta Aleatória). Devido a característica do conjunto de dados, também foram aplicadas técnicas de balanceamento de dados. Foi feito o teste com Random Under-Sampling em conjunto com o SMOTE e utilizando o SMOTE somente. Além disso, aplicou-se a normalização e engenharia de features, antes da execução dos algoritmos. Para análise dos resultados, foram empregadas métricas de precisão, revocação e acurácia. Os algoritmos Floresta Aleatória, LightGBM e XGBoost obtiveram melhores resultados em termos de precisão e revocação quando comparado com a Regressão Logística e Naive Bayes. O algoritmo Floresta Aleatória obteve 99% de precisão e revocação, assim como o LightGBM e o XGBoost. Esse resultado corrobora com o argumento de que algoritmos de ensemble possuem melhores desempenhos, já que combinam resultados de vários modelos. Em relação ao balanceamento, os resultados do SMOTE sozinho ou com o Random Under-Sampling foram similares.

Trilha: Trabalhos de Graduação

Desenvolvimento de uma Plataforma Web Didática para auxiliar o Ensino de Estrutura de Dados

Lucas F. Borges (Universidade Federal de Uberlândia); Thiago Pirola Ribeiro (Universidade Federal de Uberlândia); Ana Cláudia Martinez (Universidade Federal de Uberlândia)

lucasfernado84@gmail.com; thiagopirola@gmail.com; aclaudia.martinez@gmail.com

Resumo: Com a diversidade de tecnologias presentes hoje no dia a dia de um aluno, um grande desafio para os professores é conseguir utilizar as tecnologias disponíveis para tornar os conteúdos ministrados mais atrativos, compreensíveis e motivadores. A adoção de imagens e vídeos auxilia o professor na apresentação do conteúdo, porém se o aluno pudesse trocar os valores, conseguiria entender melhor o funcionamento do objeto de estudo e, nesse contexto o uso de simuladores computacionais auxilia muito os alunos. Pensando nos cursos da área de computação, as disciplinas iniciais de programação e algoritmos exigem um certo grau de abstração que, normalmente, não é cobrado no ensino médio. Com isso, quando os alunos são apresentados aos conceitos de laços de repetição e estruturas homogêneas (vetores), iniciam as dificuldades com as abstrações, iniciando o desinteresse pela matéria, aumentando a retenção e evasão nos cursos. Buscando esse assunto na literatura, percebeu-se que existem simuladores para muitas áreas e com diversas finalidades, além dos mais diversificados tipos de implementações e propósitos. Existem diversos simuladores para conteúdos mais complexos de Estrutura de Dados: árvores, filas e listas, porém quando o foco são estruturas mais básicas, poucos são encontrados. Dentre os diversos simuladores analisados, alguns se destacaram: Data Structure Visualizations, VisuAlgo, Data Structure Simulation e Data structure and algorithm tutorial. Nos trabalhos analisados da literatura, a maioria, demonstram que houve um ganho no aprendizado dos alunos que utilizaram os simuladores comparado aos alunos que não utilizaram. Neste trabalho, inicialmente realizou-se um estudo com professores de programação e estrutura de dados para levantar as maiores dificuldades de alunos dos cursos de programação e foi proposto o desenvolvimento de um simulador baseado em plataforma Web. Após analisar muitos simuladores disponíveis na literatura, verificando as estruturas simuladas, os softwares utilizados para o desenvolvimento e as interfaces de usuários implementadas, o simulador proposto utiliza as tecnologias: HTML 5, CSS 4, Javascript, PHP 7.4 e PostgreSQL 12.3. Diferentemente dos simuladores da literatura, o simulador proposto (debugandoED - <https://debugandoed.facom.ufu.br/>) foi pensado que o usuário tenha uma visão não somente do que ocorre visualmente a cada interação, mas consiga observar qual comando foi executado e o que aconteceu na memória. Toda parte de código exibida para o usuário está em linguagem C e, além disso, a exibição da memória ilustra para o usuário como e onde estão as variáveis. Em termos práticos, seria como executar um “debug” de uma estrutura e acompanhar os passos, sem utilizar um Ambiente de Desenvolvimento Integrado (IDE) ou programar. Já estão em fase de implementação as estruturas básicas de pilha, fila, listas, árvores e ordenação, porém ainda não estão disponibilizadas. Além disso, o DebugandoED terá versões em português e inglês.

Trilha: Trabalhos de Graduação

Análise de métodos de construção do grafo para detecção de comunidades

Lucas Gabriel T. Araújo (Universidade Federal de Uberlândia); Murillo Guimarães Carneiro (Universidade Federal de Uberlândia)

lucas.teodoro@ufu.br; mgcarneiro@ufu.br

Resumo: O agrupamento de dados consiste em separar uma base ou conjunto de dados em grupos coesos baseado em algum critério de similaridade. Tal critério frequentemente leva em conta apenas os atributos físicos dos dados utilizando, por exemplo, medidas de distância ou centróides. Na maior parte das vezes os algoritmos tratam esses dados na forma de vetor de atributos, porém também é possível encontrar grupos (ou comunidades) através de técnicas de detecção de comunidades caso esses dados estejam dispostos em forma de uma rede complexa. Redes complexas são grafos com padrões de conexão não triviais, os quais representam um conjunto de ferramentas robustas para representar e modelar sistemas reais, as quais permitem examinar, além dos atributos físicos, a estrutura topológica dos dados. Neste trabalho é considerado um framework para detecção de comunidades proposto recentemente na literatura, o qual consiste em duas maiores etapas: construção do grafo e detecção de comunidade. A primeira etapa é responsável por gerar uma representação em rede para dados dispostos em outros formatos, tais como vetor de atributo, imagem, texto, etc. A segunda é responsável por encontrar comunidades a partir da rede gerada. Motivado pelos bons resultados obtidos por tal framework, bem como nas poucas análises relacionados à construção do grafo existentes na literatura, esta pesquisa almeja investigar os principais métodos de construção do grafo no contexto desse framework de modo a analisar comparativamente os desempenhos desses métodos considerando os algoritmos de detecção de comunidades mais usados da literatura. Além do método k-vizinhos mais próximo (kNN) simétrico, avaliado no framework original, o estudo avalia outros métodos baseados na heurística kNN e na vizinhança de raio (N). Os diferentes métodos de construção do grafo foram então incorporados ao framework para detecção de comunidades e aplicados ao problema de agrupamento de padrões invariantes em imagens. Para avaliação das comunidades encontradas foram utilizadas métricas baseadas em ground-truth. Os resultados mostram que utilizar apenas o kNN-simétrico para construção do grafo não é a melhor escolha e que a combinação de alguns desses métodos, tais como N+kNN-mútuo, podem melhorar consideravelmente a qualidade dos grupos detectados pelo framework.

Trilha: Trabalhos de Graduação

Análise de vulnerabilidades de segurança usando técnicas de mineração de dados

Lucas G. Mendes (Universidade Federal de Uberlândia); Marcelo Maia (Universidade Federal de Uberlândia); Rodrigo Miani (Universidade Federal de Uberlândia)

lucas.guimaraes@ufu.br; marcelo.maia@ufu.br; miani@ufu.br

Resumo: Com os avanços tecnológicos, as vulnerabilidades se tornaram mais complexas e perigosas, tal afirmação pode ser comprovada ao observar a rápida disseminação dos vírus, que estão mais sofisticados. Uma alternativa para amenizar esse problema foi a criação do NVD (National Vulnerability Database) que é o repositório de dados de gerenciamento de vulnerabilidade do governo dos Estados Unidos. Esses dados permitem a automação desse gerenciamento, além de medidas de segurança e conformidade, como por exemplo, a correção de falhas encontradas em um determinado software, com objetivo de deixá-lo seguro e funcional. O NVD inclui dados de referências de listas de verificação relacionadas à segurança, falhas de software relacionadas a ela, configurações incorretas, nomes de produtos e métricas de impacto. Esses dados servirão de fonte para o desenvolvimento da pesquisa. Assim, obtém-se estatísticas relacionadas à nossa linha de estudos. Além disso, neste projeto iremos trabalhar com a plataforma do Stack Overflow, coletando posts que tenham alguma citação relacionada à segurança da informação, em especial às citações de CVE (Common Vulnerabilities and Exposures) e CWE (Common Weakness Enumeration). O objetivo geral deste trabalho é usar métodos de Mineração de Dados para analisar vulnerabilidades de segurança presentes na base de dados da NVD e que são citados no Stack Overflow, o qual é um site de perguntas e respostas voltado para profissionais e entusiastas na área de programação de computadores. Para alcançar tal objetivo é necessário o conhecimento em assuntos como: conceitos básicos sobre segurança da informação, ferramentas estatísticas, banco de dados, Big Data e programação. A metodologia utilizada em algumas das etapas deste trabalho envolveu o desenvolvimento de scripts, parsers e expressões regulares para a obtenção e manipulação dos dados. Além disso, também foram utilizados como fonte de informações dois bancos de dados: o da NVD e o do Stack Overflow. Os dados coletados foram salvos em formato de arquivo .csv e posteriormente analisados utilizando a ferramenta Excel e linguagens de programação como Python e R. Os resultados preliminares indicam um total de 1285 CVE's diferentes, com média de CVSS - indicador do grau de severidade da vulnerabilidade - igual a 6,3. Além disso, 74,7% dos CVE's apareceram em apenas uma postagem. Outro resultado alcançado se refere à quantidade de CVE's que possuem exploits presentes no mundo real, que totalizou 6,38% e possui média de CVSS igual a 8,14. Ademais, para analisar o impacto das postagens relacionadas às vulnerabilidades foram analisados os votos de cada uma no Stack Overflow, que obteve média de 6,22.

Trilha: Trabalhos de Graduação

Estudo da linguagem Rust na construção de aplicações embarcadas

Marcelo Braga (Universidade Federal de Uberlândia); Lukas Gomes (Universidade Federal de Uberlândia); Cecília Silva (Universidade Federal de Uberlândia); Marcelo Almeida (Universidade Federal de Uberlândia)

marcelopbj@hotmail.com; lukas.gomes@ufu.br; cecilia.carneirosilva@gmail.com;
marcelo.barros@ufu.br

Resumo: Atualmente, a linguagem C é a principal escolha dos desenvolvedores para a implementação de softwares embarcados. Apesar de ser uma linguagem que se sobrepõe em segurança e robustez, recentes falhas sugerem uma mudança de paradigma, conduzindo a busca por métodos de construção de firmwares focados na segurança da aplicação. Assim, surge a linguagem Rust como uma alternativa a esse cenário, apresentando-se promissora para substituir as linguagens convencionais em sistemas embarcados. Criada com o propósito de renovar o desenvolvimento de sistemas, Rust se apresenta como uma linguagem segura, uma vez que alia segurança, eficiência e robustez, bem como oferece um controle de baixo nível com abstrações de alto nível. Propõe-se verificar por meio deste, a possibilidade e viabilidade do uso da linguagem Rust em sistemas embarcados através de uma apresentação prática, analisando os aspectos construtivos e métricas quantitativas. Para isso, foi desenvolvido uma mesma aplicação a nível de registrador na linguagem Rust e em C, para o mesmo hardware. A aplicação foi simples, mas serve como prova de conceito, em Rust ainda existem esforços na criação de soluções mais complexas. Na construção do software da linguagem C, não foi utilizado bibliotecas de hardware desenvolvidas pelo fabricante. A garantia de segurança proposta pela linguagem Rust, é dada por meio do sistema de tipos usado para reforçar as regras de negócio na configuração dos periféricos. Os conceitos de propriedade e empréstimo da linguagem foram usados para ajudar a controlar o acesso aos registros dos periféricos, o que garante um código sem aliasing e dupla configuração. A coleta dos parâmetros para a análise das aplicações, se deu por meio do tamanho dos executáveis e o tempo de execução. Para a obtenção dos valores foram utilizadas todas as configurações de otimizações do compilador disponíveis: na linguagem C o GCC e em Rust o LLVM. Para obter o tempo de execução, foi utilizado como ferramenta de identificação, um analisador lógico, que coletou os valores de frequência. A partir de uma análise dos dados obtidos, foi possível notar o maior desempenho da linguagem Rust em tempo de execução. No entanto, na análise de tamanho de código, Rust apresentou maior tamanho que em C, o que pode ser justificado pelas diversas garantias de segurança que a linguagem oferece. Nas aplicações apresentadas, a linguagem Rust, quando comparada com a linguagem C, apresentou superioridade, como maior processamento e seus aspectos construtivos que garantem maior confiabilidade e segurança da aplicação. Concluiu-se que é possível evidenciar os benefícios que a linguagem Rust pode prover, como sua eficácia na utilização aplicada à sistemas embarcados. No entanto, os resultados obtidos não devem ser ir além da aplicação desenvolvida, pois ainda não existem outros comparativos formais de performance entre as linguagens, sendo esperado que varie de acordo com cada funcionalidade utilizada.

Trilha: Trabalhos de Graduação

Redes Neurais Artificiais para modelagem e previsão de dados meteorológicos da região do Alto Paranaíba do Triângulo Mineiro

Pryscilla R Medeiros (Instituto Federal do Triângulo Mineiro); André Batista (Instituto Federal do Triângulo Mineiro)

pryscillaromao@hotmail.com; andreluiz@iftm.edu.br

Resumo: As hidrelétricas ainda são os principais meios de geração de energia no Brasil, correspondendo a 66,1% de toda a energia produzida, segundo dados do Operador Nacional do Sistema Elétrico (ONS), em decorrência deste fato, é de suma importância a visualização da vazão antes que ocorra, pois se a máxima for estimada, os custos em manutenção se elevam e se for subestimada pode haver inundações e os danos podem trazer riscos à fauna e flora e a vida humana. Com base nisto, foram pegos dados da estação Ponte do Prata, bacia Rio Paraná, sub-bacia Rio Paranaíba, no município de Ituiutaba-MG, de 2016 até 2019 da plataforma HidroWeb da Agência Nacional das Águas(ANA). Estes dados passaram por um pré-processamento, onde foram retirados buracos no dataset e postos no formato de janela que caracteriza uma série temporal, e foram usados para treinamento e teste do modelo da rede neural. O modelo foi construído usando a linguagem de programação Python, com as bibliotecas Pytorch, Pandas e Numpy e rodados em um computador com as configurações 1TB de HD, 8GB de RAM, processador Intel i5-9300H e Windows 10 Home. A permutação dos parâmetros Learning rate (LR), momentum, primeira camada escondida e segunda camada escondida, escolhidas de forma manual, resultaram em 144 testes, e destes foram escolhidos cinco para análise, esses modelos foram escolhidos com base em seu erro pós-treinamento correspondente ao Mean Squared Error (MSE). Estas amostras passaram pelo os critérios de avaliação BIC e AIC. Analisando as cinco amostras percebe-se o porquê do das redes neurais estarem crescendo no mercado como um modelo de previsão, pois as redes neurais acompanham as altas e baixas, permitindo a visualização gráfica mais precisa, do que se comparado aos modelos probabilísticos, como por exemplo o modelo ARIMA. Pois estes modelos, apesar de terem uma melhor performance se comparadas as redes neurais artificiais, graficamente não é possível a visualização de tendências e da percepção de informações com base em por exemplo sua sazonalidade, dentre outros fatores que são possíveis visualizar por meio de uma série temporal.

Trilha: Trabalhos de Graduação

Chapter II

TRABALHOS DE MESTRADO

Imagens infravermelhas e redes convolucionais para detecção de câncer de mama

Caroline Barcelos Gonçalves (Universidade Federal de Uberlândia); Jefferson Souza (Universidade Federal de Uberlândia); Henrique C Fernandes (Universidade Federal de Uberlândia)

barceloscarol.cbg@gmail.com; jrsouza@ufu.br; henrique.fernandes@ufu.br

Resumo: O câncer de mama é uma doença que afeta muitas mulheres em todo o mundo, sendo o segundo mais comum e o segundo causador de mortes em mulheres. Sabe-se ainda que quando descoberto precocemente a paciente tem boas chances de recuperação portanto a detecção precoce da doença é um problema relevante. Neste trabalho, redes convolucionais foram usadas para classificar pacientes com e sem câncer de mama em imagens infravermelhas. Foram consideradas as imagens estáticas da base de dados pública DMR totalizando 188 imagens de cada classe que correspondem a 38 pacientes. Para a classificação fez-se uso de algumas redes convolucionais do estado da arte, VGG-16, Resnet50 e Densenet201, que apresentaram bons resultados na literatura utilizando 'transferência de aprendizado' alterando-se apenas a arquitetura da camada densamente conectada. Nas redes foram avaliadas diferentes arquiteturas para a camada densamente conectada, considerando o otimizador Adam para todos os testes, duas learning rates foram testadas 0.001 e 0.0001 e aplicou-se o early stopping para evitar overfitting. As redes foram avaliadas utilizando apenas as 188 imagens originais de cada classe e ainda utilizando aumento de dados. Para o aumento de dados consideramos: rotação (com ângulo entre -30 a 30) e flips horizontais e verticais. Foi necessário ainda um pré-processamento para que as imagens pudessem ser utilizadas nas redes pré treinadas e para isto utilizou-se normalização min max com replicação de canais. Os melhores resultados obtidos para o conjunto de teste são apresentados na sequência. Para os experimentos sem aumento de dados para a VGG foram 88% de acurácia, 0.88 F1-score, 96% de sensibilidade, e 80% para especificidade com a configuração de taxa de aprendizado 0.001 e 2 camadas na parte densamente conectada da rede com respectivamente 512 e 2 neurônios de saída. Para a densenet obteve-se 84% de acurácia, 0.84 de F1-score, 82% de sensibilidade e 86% de especificidade considerando learning rate 0.0001 e 2 camadas na parte densamente conectada com (1024, 2). Para a resnet a mesma configuração obteve os melhores resultados sendo eles 79% de acurácia, 0.7 de F1-score, 81% de sensibilidade e 89% de especificidade. Os resultados com aumento de dados tiveram resultados inferiores aos experimentos sem aumento de dados obtendo-se os melhores resultados com a Resnet com learning rate de 0.001, e a arquitetura densamente conectada com 3 camadas contendo (4096, 4096, 2) neurônios de saída com 76% de acurácia, 0.76 para F1-score, 75% de sensibilidade e 77% de especificidade. As demais redes não apresentaram resultados acima de 0.7 em todas as métricas consideradas e, portanto, não são reportadas. As imagens estáticas devido a presença de imagens rotacionadas onde apenas uma das mamas é visível podem impactar negativamente na rede e é um problema que pretendemos tratar posteriormente. Mesmo com esse ponto de atenção os resultados obtidos inicialmente se demonstram promissores.

Trilha: Trabalhos de Mestrado

Uma proposta para implementação prática de auto-configuração especializada de elementos de borda em redes SDN Utilizando a arquitetura SONAr

Daniel Ricardo C. Oliveira (Universidade Federal de Uberlândia); Maurício Amaral Gonçalves (Universidade Federal de Uberlândia); Natal Vieira Souza Neto (Universidade Federal de Uberlândia); Flavio Oliveira Silva (Universidade Federal de Uberlândia); Pedro Rosa (Universidade Federal de Uberlândia)

drcoliveira@gmail.com; mauricioamaralg@ufu.br; natalneto@ufu.br; flavio@ufu.br; pfrosi@ufu.br

Resumo: A crescente complexidade das redes de computadores para suportar as atuais aplicações vêm sendo percebida como um fator limitante em termos técnicos e econômicos para sua evolução. Uma resposta a essa questão é a proposta de novas arquiteturas de redes autônomas, baseadas em Software-Defined Networks (SDN), com implementação de funcionalidades chamadas self-* e, entre essas, a self-configuration (auto-configuração). Nesta dissertação, o objetivo é a implementação de um caso-modelo de self-configuration especializado de redes: o sistema deverá que reconhecer um novo elemento e configurar um determinado segmento de rede de modo que o mesmo funcione em um modelo “plug-and-play”. A abordagem será baseada na arquitetura Self-Organizing Networks Architecture (SONAr), em desenvolvimento por um grupo de pesquisadores da FACOM/UFU. Há poucos trabalhos práticos nesse sentido: as funcionalidades Self-Organized Network (SON) estão mais maduras no contexto de redes móveis, porém em um nível restrito aos protocolos específicos 3GPP. O framework SELFNET também foi definido como uma tentativa de incorporar o SON em um contexto SDN/NFV, porém não contempla a funcionalidade de self-configuration. Outras propostas práticas de self-configuration já foram abordadas em outros trabalhos, mas esses apresentam limitações em relação ao tamanho e complexidade da rede, ou são destinadas a finalidades mais específicas. O SONAr já apresentava uma entidade de self-configuration com módulos de bootstrap e plug-and-play. O trabalho aqui apresentado incorpora um sistema de plug-and-play para elementos que requerem um alto grau de especificidade em sua configuração na rede, como estações rádio-base de redes móveis. Isso faz com que a aplicação do SONAr não restrinja-se a um caso geral, mas suporte quaisquer modelos de configuração utilizados em redes reais e torne-se uma opção viável dentro do contexto da indústria de telecomunicações, sobretudo quando pensa-se na futura implementação de milhares de elementos 5G. A implementação do protótipo já encontra-se finalizada e em funcionamento, sendo que o processo de specialized self-configuration converge em poucos segundos, mesmo com um grande número de elementos a serem configurados.

Trilha: Trabalhos de Mestrado

Dispositivo de monitoramento e acompanhamento de frequência cardíaca e temperatura corporal

Fernanda Rodrigues Oliveira (Universidade Federal de Uberlândia); Shiguo Nomura (Universidade Federal de Uberlândia)

fernandaroliveira@hotmail.com; shiguo.nomura@hotmail.com

Resumo: No Brasil, várias mudanças têm ocorrido nas necessidades de saúde devido, dentre outros, ao aumento da expectativa de vida da população para 80 anos até 2025, melhoria nas condições de vida e queda nas taxas de natalidade. O envelhecimento da população, o aumento das doenças crônico-degenerativas e suas complicações, acidentes automobilísticos e problemas sociais são fatores que contribuem para o aumento das práticas de cuidados em saúde no domicílio, denominada atenção domiciliar (AD). Porém, esta modalidade de atendimento é acompanhada de vários desafios, dentre eles, a não disponibilização de tecnologias de informação e comunicação (TICs) adequadas. Considerando a área de estudo, e-Saúde, que origina da aplicação das TICs nos cuidados à saúde, a popularização do acesso à Internet e a agregação deste acesso por meio de smartphones, pesquisadores e empresas de todo o mundo perceberam que a Internet poderia ser acessada por outros dispositivos inteligentes, levando ao conceito de Internet das Coisas. Neste contexto, o objetivo deste trabalho foi desenvolver um dispositivo de monitoramento de frequência cardíaca e temperatura corporal, dada a importância destes sinais vitais, podendo este ser um recurso de apoio ao cuidado de idosos ou mesmo outros pacientes em AD e ainda amenizar as consequências da distância física do profissional de saúde e/ou familiar/cuidador. Muitas aplicações de e-Saúde foram desenvolvidas como o aplicativo Personal Digital Assistant (PDA) da Universidade Federal de São Paulo. O PDA é um protótipo para coleta e análise de informações em saúde de maneira organizada e controlada através de dispositivos móveis para o atendimento ambulatorial das Unidades Básicas de Saúde de Embu-SP. O dispositivo desenvolvido neste trabalho é composto pelo microcontrolador ESP8266, o sensor de frequência cardíaca da World Famous Electronics llc. e o sensor de temperatura DS18B20. Para realizar a comunicação entre o hardware e aplicativo do usuário foi utilizado o Blynk App versão gratuita. O sistema desenvolvido é de custo bastante acessível e encontra-se registrado sob o número 512019002793-0 no INPI. O diferencial deste sistema em relação às propostas da literatura, como a citada acima, é a capacidade de comunicação via Wi-Fi, possibilidade de coleta de dados clínicos do paciente em AD e envio em tempo real via Internet para a equipe médica. Assim, estes profissionais podem orientar cuidadores a agirem para reverter o quadro clínico do paciente ou até para intervir rápido em casos de necessidade de socorro imediato. Portanto, o monitoramento à distância pelo sistema proporcionaria agilidade no acesso às informações de um paciente, bem como reduziria o tempo de resposta da equipe para os cuidadores/familiares deste paciente. As consequências da implantação do sistema seriam a melhoria na qualidade dos serviços por redução da sobrecarga dos profissionais de saúde e o aumento da confiança do paciente em AD através de um cuidado mais eficiente.

Trilha: Trabalhos de Mestrado

Detecção de Intrusão sobre Pacotes de Redes utilizando Algoritmos de Fluxos Contínuos de Dados

Gilberto Olimpio Júnior (Universidade Federal de Uberlândia), Elaine Ribeiro de Faria Paiva (Universidade Federal de Uberlândia)

gilberto.olimpio@gmail.com; elaine@ufu.br

Resumo: À medida que a importância dos dados aumenta, há uma necessidade de protegê-los na infraestrutura que eles vivem e transitam. Cada vez mais, a tentativa de acesso não autorizado a esses dados torna-se um problema de segurança da informação. Sistemas de Detecção de Intrusão (ou IDS - Intrusion Detection Systems em inglês) tentam identificar e prevenir essas tentativas de ganhar acesso não autorizado aos dados por meio das redes de computadores, executando a inspeção de pacotes individualmente ou na forma de fluxos desses pacotes. Considera-se o trabalho de um IDS como uma tarefa de classificação sobre fluxo dos pacotes gerados continuamente com uma distribuição de dados não-estacionária. Para lidar com esta tarefa, há a necessidade de se utilizar algoritmos de classificação de fluxo contínuos de dados de última geração que, de fato, aprendem de forma contínua o que é um tráfego normal e o que pode ser um ataque. O objetivo é verificar se o uso de dados dos pacotes apresenta desempenho semelhante à inspeção de dados dos fluxos em IDS. Assim, utilizando somente um subconjunto dos dados do cabeçalho dos pacotes dos Protocolos IP (Internet Protocol), TCP (Transfer Control Protocol) e UDP (User Datagram Protocol) deseja-se verificar se ele é suficiente para que a tarefa de classificação de pacotes maliciosos seja bem-sucedida. Foram selecionados seis algoritmos de classificação de fluxos contínuos de dados sendo que, cinco deles utilizam a técnica de comitê de classificadores: Accuracy Updated Ensemble, Leveraging Bagging, Limited Attributes Classifier, Oza Bagging ADWIN e Oza Bagging ASHT e um utiliza um único classificador Single Classifier Drift. Houve uma comparação entre esses algoritmos a fim de identificar qual obteve o melhor desempenho para a classificação de pacotes maliciosos resultando assim, na identificação do melhor algoritmo para esse tipo de tarefa. Os resultados são demonstrados por meio de experimentos utilizando o conjunto de dados CICIDS2017 que cumpre um requisito importante: contém os dados de ataques recentes e disponibiliza para os experimentos tanto os dados dos pacotes brutos quanto dos fluxos gerados por esses pacotes. Para a análise dos resultados foram utilizadas as medidas disponíveis na literatura e testes estatísticos para validar todos os resultados apresentados nos experimentos. Os resultados obtidos mostraram que o desempenho da classificação do fluxo de pacotes quanto dos cabeçalhos dos pacotes é semelhante, não existindo diferença estatística significativa. Entre os seis classificadores o Leveraging Bag obteve o melhor desempenho nas diferentes medidas de avaliação. Um IDS baseado em cabeçalhos usando dados em nível de pacotes possui o desempenho de detecção equivalente àqueles em nível de fluxo indicando assim, que é possível criar mecanismos de decisão para operar ao nível de pacote fornecendo resposta imediatas de forma satisfatória para a proteção das redes de computadores.

Trilha: Trabalhos de Mestrado

Nova heurística de fluxo de informação para classificação de alto nível usando redes complexas

Janayna Moura Fernandes (Universidade Federal de Uberlândia); Murillo Guimarães Carneiro (Universidade Federal de Uberlândia)

fernandesmjnanayna@gmail.com; mgcarneiro@ufu.br

Resumo: Redes Complexas são grafos com padrões de conexões não triviais. A internet, as redes sociais, o cérebro e a bolsa de valores são exemplos de redes complexas. Técnicas baseadas em grafos podem, além de considerar aspectos do espaço físico como distância, proximidade, distribuição, considerar aspectos funcionais, topológicos, estruturais. A classificação de dados é uma das tarefas mais relevantes da literatura e amplamente investigada no aprendizado de máquina. Nesse sentido, o uso de redes complexas permite, além de examinar atributos físicos dos dados, também considerar informações topológicas a partir da configuração deles em uma rede. Diferentes métodos da literatura baseados em redes complexas, tais como conformidade de padrão e caracterização de importância, demonstram que tais informações permitem um melhor desempenho preditivo na detecção do padrão semântico dos dados. Neste trabalho investiga-se a técnica de caracterização de importância devido a sua menor complexidade de parâmetros e a unificação da análise de associações físicas e topológicas em um único algoritmo. A técnica original é representada por duas fases principais: a fase de treino, onde a rede é construída a partir dos dados de entrada (na forma de vetor de atributos) e posteriormente são calculadas uma medida de eficiência no fluxo da informação e outra de importância baseada no PageRank; e a fase de teste consiste na inserção virtual de um dado objeto de teste na rede baseado na melhoria da eficiência no fluxo de informação e a atribuição de importância ao objeto de teste, de modo que o mesmo seja atribuído à classe na qual recebeu maior valor de importância. Nesse sentido, a contribuição deste trabalho é o desenvolvimento de uma nova heurística de inserção baseada em menores caminhos médios, bem como investigar diferentes medidas de centralidade para caracterizar a importância dos vértices, tais como grau médio e proximidade. Experimentos preliminares conduzidos em seis bases artificiais e em dez bases reais mostraram que essas novas estratégias podem aprimorar a extração de informação subjacente nos dados.

Trilha: Trabalhos de Mestrado

Algoritmos evolutivos multiobjetivo e modelo ETC para escalonamento de tarefas em ambientes distribuídos

José Junio Ribeiro de Sousa (Universidade Federal de Uberlândia); Paulo Henrique Ribeiro Gabriel (Universidade Federal de Uberlândia)

josejuniosrd@gmail.com; phrg@ufu.br

Resumo: Problemas de escalonamento de tarefas aparecem em diversas aplicações do mundo real. Esses problemas consistem em atribuir diversas tarefas sobre um conjunto limitado de máquinas de modo a otimizar um ou mais critérios de desempenho. Problemas de escalonamento aparecem comumente na indústria, quando se deseja diminuir a ociosidade das máquinas a fim de atingir o máximo de produtividade. Já em computação, busca-se atribuir aplicações paralelas a recursos computacionais como, por exemplo, CPU. Diversos algoritmos têm sido propostos a fim de lidar com esse problema. Neste trabalho, iremos abordar o cenário de tarefas independentes em um ambiente computacional heterogêneo. Além disso, busca-se minimizar dois critérios de otimização: makespan e flowtime. O primeiro consiste no tempo total de processamento das máquinas; quando menor o makespan, mais eficiente é a solução encontrada. O segundo, por sua vez, representa o somatório do tempo de processamento das tarefas; sua minimização reduz a média de tempo de execução e maximiza a produtividade. Os conjuntos de tarefas e de recursos podem ser representados de diversas formas. Nesse trabalho, utiliza-se o modelo denominado Expected Time to Compute (ETC), sendo definido como a expectativa do tempo de processamento de cada tarefa em cada processador. Esse mapeamento é feito utilizando uma tabela (ou matriz) em que as tarefas são indexadas nas linhas e os recursos nas colunas. Como ferramenta de otimização, serão empregados algoritmos evolutivos (AEs). Os AEs são métodos de busca e otimização inspirados nos mecanismos de evolução dos seres vivos. Esses algoritmos procuram imitar a seleção natural e sobrevivência do mais apto; assim, geram uma população aleatória de indivíduos e através do processo evolutivo, a cada iteração os indivíduos mais capazes geram uma nova população. Cada iteração é denominada geração, e a cada geração tende a ocorrer uma evolução. Este projeto propõe o desenvolvimento de algoritmo evolutivo multiobjetivo, abordando os critérios de otimização já mencionados e tendo como base o modelo ETC. Esta abordagem permitirá a obtenção de novas soluções de escalonamento quando estamos em um cenário multiobjetivo e em ambiente distribuído.

Trilha: Trabalhos de Mestrado

Detecção e correção de distorções não lineares para co-registro multiespectral de imagens agrícolas obtidas por veículos aéreos não tripulados

Leandro Silva (Universidade Federal de Uberlândia); Jocival Dantas Dias Júnior (Universidade Federal de Uberlândia); Joao F. Mari (Universidade Federal de Viçosa); Mauricio Escarpinati (Universidade Federal de Uberlândia); André R. Backes (Universidade Federal de Uberlândia)

leandrofurtado.ufv@gmail.com; jocival.dias@ufu.br; joaofmari@gmail.com; mauricio@ufu.br; arbackes@yahoo.com.br

Resumo: No último meio século, o crescimento populacional e a demanda por alimentos impulsionaram o uso, cada vez maior, de recursos tecnológicos no que chamamos de Agricultura de Precisão (AP). Dentre essas tecnologias, a utilização de Veículos Aéreos Não Tripulados (VANTs) tem-se destacado por auxiliar, potencializar e otimizar a produção agrícola. As imagens capturadas por VANTs permitem uma visão detalhada da região analisada, visto que voo ocorre em altitudes baixas e médias (50m a 400m). Além disso, existe uma grande variedade de sensores (câmeras RGB, sensores de captura de calor, câmeras multi e hiperespectrais, entre outros), cada um com características próprias e capaz de produzir informações que exigem diferentes tipos de análises. O processo de aquisição das imagens ocorre sequencialmente e, posteriormente, ao término do voo, as imagens são organizadas em um mosaico, o qual representa visualmente toda a área. Devido à natureza do voo, dois tipos principais de deformações entre espectros podem ser observadas, são elas: (1) lineares e (2) não lineares. As deformações lineares consistem, por exemplo, em transformações de rotação, translação e escala; já nas não lineares possuem características variadas e sem um padrão previamente definido. Para a correta análise dessas imagens é necessário identificar e corrigir essas distorções, de modo a se extrair o máximo de informação relevante e capaz de auxiliar as decisões e atividades inerentes da AP. Devido às características supracitadas das distorções não lineares, técnicas tradicionais de processamento digital de imagens podem não ser suficientes para identificar e corrigir esse tipo de deformação. Essas deformações podem ser descritas através de campos vetoriais, os quais determinam o quanto um determinado ponto foi deslocado. Diante disso, propõe-se o treinamento progressivo de redes neurais profundas para as tarefas de identificação da presença da distorção e, caso ela exista, a predição dos campos vetoriais de deformações. Resultados preliminares evidenciam que tais técnicas se mostram robustas para solucionar tais problemas. Além disso, para a identificação da presença (ou não) de deformação, há resultados consolidados no presente trabalho que demonstram que a utilização de uma arquitetura de rede neural convolucional consideravelmente menor, comparada com técnicas tradicionais da literatura, em relação ao custo computacional, consegue solucionar o problema. Esse tipo de resultado sugere a possibilidade de utilização da técnica proposta em um sistema embarcado ao VANT, onde as imagens são marcadas automaticamente quanto a presença (ou não) de deformação, fomentando, por consequência, o processo de estimar os campos vetoriais de deslocamento.

Trilha: Trabalhos de Mestrado

Aprendizado de padrões em sinais EEG para apoio ao prognóstico de pacientes em coma

Lucas Zago Bissaro (Universidade Federal de Uberlândia); Camila Davi Ramos (Universidade Federal de Uberlândia); João Batista Destro Filho (Universidade Federal de Uberlândia); Murillo Guimarães Carneiro (Universidade Federal de Uberlândia)

lucasbissaro@gmail.com; camiladavi.r@gmail.com; jbdestrof@yahoo.com; mgcarneiro@ufu.br

Resumo: O exame de eletroencefalografia (EEG) monitora e grava as atividades elétricas do cérebro. Entre as suas principais vantagens destaca-se o seu baixo custo e a sua natureza não invasiva, o que o torna adequado à realidade da maioria dos hospitais públicos brasileiros. A análise do EEG é realizada por um médico, que procura por padrões comuns para determinados prognósticos. Um problema é que um exame de 20 minutos, pode gerar mais de 200 mil pontos, o que torna a análise desses padrões um trabalho exaustivo e não eficiente. Nesse sentido, a popularização dos algoritmos de aprendizado de máquina tem contribuído para o aprendizado automático de padrões no EEG para prognóstico de algumas doenças, tais como Parkinson, Alzheimer, esquizofrenia, epilepsia e afasia. A análise de Comatosos por aprendizado de máquina é um problema pouco explorado na literatura. As abordagens costumam ser mais tradicionais, majoritariamente baseando-se em procedimentos de extração de atributos dos sinais e análises estatísticas convencionais. As poucas abordagens voltadas para modelagem preditiva do coma, utilizam o exame fMRI e normalmente consideram dois grupos de desfecho: resultado favorável ou desfavorável. O grupo favorável diz respeito a pacientes com bons resultados neurológicos como recuperação completa ou incapacidade moderada, enquanto o grupo desfavorável representa incapacidade severa, vigília sem resposta (estado vegetativo) ou morte. Aqui estendemos a literatura, adicionando um novo grupo chamado de morte cerebral. De acordo com a lei brasileira, morte cerebral é uma condição clínica de coma irreversível de causa conhecida com ausência de reflexos no tronco cerebral e apneia. Portanto, estamos interessados no desenvolvimento e aplicação de técnicas de classificação de dados, para permitir uma detecção precoce em pacientes mais propensos a morte cerebral. O prognóstico precoce pode melhorar a quantidade e a qualidade dos órgãos doados, pois uma variedade de alterações fisiológicas ocorrem durante e após o processo de morte cerebral, tais como alterações inflamatórias e hormonais generalizadas, mudanças agudas, entre outras, que podem afetar a função do órgão doador e aumentar sua propensão à rejeição. Neste trabalho, investigamos métodos de aprendizado sequencial e profundo para apoio ao prognóstico de pacientes comatosos. O objetivo é que, diferente da maior parte das abordagens consideradas na literatura, as soluções desenvolvidas sejam capazes de diminuir a necessidade do médico na etapa de seleção e preparação dos dados. Resultados preliminares baseados na análise de padrões sequenciais a partir de técnicas de computação de reservatório indicam o caráter promissor da pesquisa. Por exemplo, considerando apenas os sinais capturados por um dos vinte eletrodos EEG, a acurácia média alcançada foi de 78%. Espera-se que o trabalho possa contribuir para melhorar a gestão e a qualidade do tratamento de pacientes em coma.

Trilha: Trabalhos de Mestrado

Processamento e Classificação de Séries Temporais Aplicados ao Desafio PLAsTiCC

Márcio Antônio F. Júnior (Universidade Federal de Uberlândia); Marcelo Albertini (Universidade Federal de Uberlândia);

marcio55@ufu.br; albertini@ufu.br

Resumo: No deserto do Atacama, em Cerro Pachón - Chile, está sendo construído o Rubin Observatory. Este observatório é constituído pelo Simonyi Survey Telescope de 8,4 metros, a maior câmera digital do mundo, um complexo sistema de processamento de dados e uma plataforma educacional. Um de seus projetos, a pesquisa de 10 anos Legacy Survey of Space and Time (LSST), foi desenvolvido para explorar o Transient Optical Sky. Através de 20TB de dados por noite e observações que cobrem mais da metade do céu em poucas noites, o LSST irá revolucionar nosso entendimento sobre o céu que está em constantes mudanças. A fim de desenvolver o sistema que irá processar todos esses dados, foi criado o desafio PLAsTiCC com o seguinte questionamento: “Quão bem podemos classificar objetos dos céus que variam em brilho a partir de séries temporais simuladas do LSST, com todos os seus desafios de não representatividade?”. O desafio PLAsTiCC disponibiliza o dataset contendo as séries temporais para ser utilizado no modelo de cada competidor e também avalia as classificações feitas por cada um deles. A competição está encerrada mas ainda é possível acessá-la. No desafio, dentre os modelos que obtiveram melhores resultados, dois classificadores se destacaram em conjunto com algumas técnicas de pré processamento: Random Forests e Redes Neurais Recorrentes para os classificadores; Processo Gaussiano para regressão e normalização das séries com valores nulos; Redução de dimensionalidade e extração de features; Resampling para gerar novos dados e aumentar a similaridade dos conjuntos de treino e de teste; Estes métodos foram utilizados principalmente para solucionar as maiores dificuldades do desafio PLAsTiCC. Essa dificuldades aparecem tanto da natureza do problema, como as características que melhor representa cada classe astronômica, tanto para simular os reais problemas que serão encontrados quando o telescópio estiver em ação, inseridos propositalmente no conjunto de dados como ruídos na amostragem e limitações físicas do observatório. Dentre elas, destacam-se: Amostragem das séries temporais desiguais: Frequência inconstante de observações; Séries com início e fim distintos; Grande diferença entre dados de treino e teste: Distribuições divergentes, conjunto de treino enviesado; Conjunto de treino pequeno em relação ao de teste; Característica importante no treino é pouco encontrada no teste; Presença de uma classe a mais no conjunto de teste; Extrair features a partir de conhecimento específico de astronomia para cada classe; Os objetivos dessa pesquisa são: solucionar os principais problemas encontrados para a classificação dos dados do PLAsTiCC; comparar os métodos mais utilizados na competição dentre os melhores colocados com outros métodos do estado da arte. Para isso, serão usados os seguintes algoritmos: Redes neurais, LGBM, Rocket, se possível o Hive-COTE, Processo Gaussiano, Kernel density estimation, Meta Mining Ontology, NAS, Spearmint e GPyOpt.

Trilha: Trabalhos de Mestrado

Um Novo Algoritmo De Difusão De Distância Para Um Modelo De Planejamento De Caminho Baseado Em Autômatos Celulares

Samuel Nametala (Universidade Federal de Uberlândia); Gina M. B. Oliveira (Universidade Federal de Uberlândia)

samuel.nametala@gmail.com; gina@ufu.br

Resumo: Um dos principais desafios da robótica, e que vem sendo foco de várias pesquisas, é encontrar um sistema, capaz de realizar o planejamento de caminho livre de colisão para robôs autônomos. Mapas de rotas, decomposição celular e campo potencial são alguns métodos tradicionais utilizados para resolver este problema, contudo recentemente, técnicas de computação Bioinspiradas estão sendo estudadas, por exemplo, o uso de Autômatos Celulares (AC). Um modelo aprimorado baseado nos trabalhos de Behring e Oliveira utilizando regras de AC, é proposto e avaliado para o planejamento de caminhos em robôs autônomos. O objetivo é construir um caminho curto e livre de colisão da posição inicial do robô até o alvo, mas diferente dos modelos anteriores onde o objetivo principal era melhorar a rota planejada, o nosso é minimizar a ocorrência de erros, relacionados com a odometria, reduzindo a distância percorrida, e tentando evitar curvas desnecessárias tanto quanto possível, para isso propomos diferenciar movimentos diagonais e cardeais e aplicar uma nova regra que corrige regiões côncavas de obstáculos. As regras de AC são usadas para: ampliar as paredes e os obstáculos internos, para evitar colisões durante a navegação do robô; provocar um aumento adicional nos obstáculos internos que possuem concavidades, para evitar que o robô entre em uma armadilha, que são regiões que levam o robô a fazer movimentos desnecessários ou provocar um risco de colisão; e espalhar a distância do objetivo dando um valor para cada célula livre, que corresponde ao quão distante ela está deste objetivo. A rota do robô é planejada a partir das informações da distância de cada célula livre até o objetivo. Experimentos foram realizados para confirmar a eficiência das novas técnicas empregadas. Foi possível observar que o novo modelo consegue retornar distâncias curtas e com trajetórias mais refinadas, graças a sua menor quantidade de giros. E as simulações no Webots mostraram que o modelo proposto é capaz de calcular rotas que podem ser percorridas com sucesso usando um robô e-puck.

Trilha: Trabalhos de Mestrado

Comparação de descritores morfológicos e não morfológicos na classificação multi classe de imagens histológicas de linfomas não Hodgkins

Tiago P. de Faria (Universidade Federal de Uberlândia); Luiz G. A. Martins (Universidade Federal de Uberlândia)

tiagofaria@ufu.br; lgamartins@ufu.br

Resumo: Um dos tipos mais comuns de câncer é o linfoma. Esta variante da doença acomete células do sistema linfático, e possui mais de 38 variantes, que podem ser divididas entre linfomas Hodgkin e linfomas não Hodgkin. O segundo tipo é o mais comum, com cerca de 90% dos casos. Para realizar um tratamento efetivo da doença, é necessário saber, além do tipo, qual a variante da doença. Porém, mesmo para médicos especialistas da área, esse diagnóstico é uma tarefa difícil. Assim, sistemas computacionais que auxiliam nesta tarefa podem ser de grande importância. Sistemas de apoio ao diagnóstico (CAD - computer aided diagnosis system) baseados em algoritmos de visão computacional e aprendizado de máquina podem ser usados para, dado uma imagem histológica do paciente, classificar o tipo de linfoma presente. Esses sistemas tem como objetivo diminuir a variabilidade dos diagnósticos e automatizar este processo que é custoso quando feito por profissionais da saúde. Este trabalho estuda técnicas para classificação multiclasse de imagens, e suas aplicações na tarefa de classificação de imagens histológicas de pacientes de linfomas do tipo não Hodgkin. A base de dados utilizada para os testes contém imagens coloridas de tecidos doentes de três variantes da doença: Mantle Cell Lymphoma (MCL), Follicular Lymphoma (FL) e Chronic Lymphocytic Leukemia (CLL). Em nossos experimentos, utilizamos descritores de imagens morfológicos e não morfológicos. O primeiro tipo, é criado com base em características geométricas dos núcleos celulares encontrados nas imagens, enquanto o segundo é extraído com base na intensidade de luz presente nesses núcleos, tanto dos 3 canais do sistema RGB, quanto da versão preto e branco da imagem. Foi avaliado o uso exclusivo de cada um dos tipos de descritores, bem quanto o uso dos dois tipos combinados. Para classificação das imagens, avaliamos o uso de 3 algoritmos clássicos de classificação: Suport Vector Machine (SVM), Decision Tree (DT) e Random Forest (RF). Com os resultados, conseguimos observar que, dentre os algoritmos utilizados, o Random Forest obteve a melhor taxa de acurácia, e os descritores não morfológicos foram os melhores para a tarefa de classificação. Mostramos ainda que a adição dos descritores morfológicos a base de dados não contribui com a tarefa de classificação, na nossa metodologia, e em alguns casos, diminuiu a taxa de acurácia dos classificadores nos nossos experimentos. Assim, definimos uma metodologia simples, e ainda efetiva para classificação das imagens, criando uma base para futuros trabalhos.

Trilha: Trabalhos de Mestrado

Chapter III

TRABALHOS DE DOUTORADO

Michelzinho: Jogo sério para o ensino de habilidades emocionais em pessoas com autismo ou deficiência intelectual

Adilmar Coelho Dantas (Universidade Federal de Uberlândia); Marcelo Zanchetta do Nascimento (Universidade Federal de Uberlândia)

akanehar@gmail.com; marcelo.zanchetta@gmail.com

Resumo: Esse trabalho apresenta um jogo sério para auxiliar no ensino de competências emocionais e sociais em pessoas com Transtorno do Espectro Autista (TEA) e deficiência intelectual. A aplicação permite desenvolver as habilidades de reconhecer e expressar as emoções básicas: alegria, tristeza, raiva, desgosto, surpresa e medo. Por meio da teoria de sistema de codificação de ação facial e técnicas de processamento digital de imagens foi possível detectar as expressões faciais e classifica-las em uma das seis emoções básicas em tempo real. Para a classificação das emoções foi desenvolvido um classificador utilizando árvore de decisão do tipo ID3 que foi implementada em uma arquitetura em nuvem para permitir maior escalabilidade e disponibilidade para o aplicação desenvolvida. Os personagens e elementos gráficos do jogo foram modelados em 3D por meio de ferramentas específicas com o objetivo de oferecer maior fidelidade das expressões faciais para cada uma das emoções básicas trabalhadas na aplicação. Para o desenvolvimento dos cenários foi realizado um estudo com profissionais da área da psicologia e da saúde para a construção de enredos e situações comportamentais e sociais para a produção correta das emoções nos usuários. Experimentos iniciais foram realizados para a aquisição e dados iniciais que serão utilizados para treinar o modelo e fornecer dados para o desenvolvimento de uma nova arquitetura para a classificação das emoções utilizando rede neural convolucional e além disso a realização de melhorias na aplicação antes de sua finalização definitiva. Essa proposta foi avaliada em imagens da base Ck+, obtendo uma taxa de precisão de 88% para a classificação das emoções básicas. Experimentos com portadores do TEA foram realizados para avaliar as contribuições e limitações do trabalho proposto, os resultados iniciais se demonstraram promissores capaz de auxiliar esses indivíduos a desenvolver e aperfeiçoar essas habilidades por meio das intervenções realizadas com a utilização do jogo.

Trilha: Trabalhos de Doutorado

Classificação de Lesões Pré Cancerosas em Imagens Histológicas da Cavidade Oral

Adriano B. Silva (Universidade Federal de Uberlândia); Marcelo Zanchetta do Nascimento (Universidade Federal de Uberlândia); Tháina A. A. Tosta (Universidade Federal do ABC); Paulo Rogério de Faria (Universidade Federal de Uberlândia); Leandro Alves Neves (Universidade de São Paulo); Alessandro Santana Martins (Instituto Federal do Triângulo Mineiro)

adrianobs@gmail.com; marcelo.zanchetta@gmail.com; tosta.thaina@gmail.com;
paulorfaria1976@gmail.com; neves.leandro@gmail.com; alessandrosm2006@yahoo.com.br

Resumo: O câncer da cavidade oral é a sexta maior causa de morte por câncer no mundo. Uma forma de combater essa doença é através do seu prognóstico, ou seja, a identificação de lesões que podem se transformar em câncer. Dentre essas lesões, as displasias são lesões pré câncer que apresentam alterações celulares como crescimento anormal e deformações na estrutura do núcleo e tecido epitelial. O diagnóstico dessa lesão normalmente é realizado por meio da análise de amostras de tecidos coradas com hematoxilina e eosina. Os sistemas de apoio ao diagnóstico são ferramentas computacionais que permitem auxiliar especialistas na identificação desse tipo de lesões. Esses sistemas geralmente envolvem etapas de segmentação de objetos, extração de características e classificação de tecidos. Dentre essas etapas, a segmentação é importante para detectar a área de objetos de interesse, permitindo a extração de características morfológicas e não morfológicas. Este trabalho apresenta uma técnica para a segmentação de núcleos e classificação de tecidos histológicos. Para a segmentação, a arquitetura de rede Mask R-CNN foi treinada para aprender as características representativas de núcleos celulares. Uma vez os núcleos segmentados, foi realizada a extração de características morfológicas e de textura. As características desses núcleos foram empregadas em um classificador polinomial para diferenciar os tecidos. A métrica de acurácia foi utilizada para avaliar o desempenho da segmentação e a medida de área sob a curva ROC (AUC) foi utilizada para avaliar a classificação. A base de dados utilizada é composta de 296 imagens de línguas de 30 camundongos submetidos a um agente carcinógeno. A etapa de segmentação apresentou valores médios de acurácia de 89,52%. Na classificação, o método forneceu uma AUC média de 0,92. Na literatura, existem poucos estudos que realizam o diagnóstico de displasias da cavidade bucal e o uso da metodologia apresentada neste trabalho contribuiu para o estado da arte e possibilita usá-la como uma ferramenta para auxiliar especialistas no estudo dessa lesão.

Trilha: Trabalhos de Doutorado

Mineração de Soluções Relevantes para Tarefas de Programação a partir de Resultados de Serviços de Busca

Adriano M. Rocha (Universidade Federal de Uberlândia); Marcelo Maia (Universidade Federal de Uberlândia)

adriano.comp2@gmail.com; marcelo.almeida.maia@gmail.com

Resumo: O desenvolvimento de software é uma atividade intensiva em conhecimento, e requer que o desenvolvedor se mantenha atualizado com novas APIs que são lançadas frequentemente. A existência de diferentes linguagens de programação e APIs a serem estudadas, é um problema recorrente, além da dificuldade por parte do desenvolvedor em encontrar um ponto de partida para começar a aprender as mesmas. Uma alternativa é buscar conhecimento na Internet, mas, devido à grande oferta de informações nesse meio, existem dificuldades em encontrar aquelas realmente úteis. Sites específicos, tais como, o Stack Overflow, concentra informações relacionadas ao desenvolvimento de software, porém, mesmo sendo específico, existem mais de 18,6 milhões de questões, ele também não cobre todas as tecnologias, além de não possuir as melhores soluções. Para buscar conteúdo em outros sites, desenvolvedores recorrem a serviços de busca (Google, Yahoo, Bing, etc.), mas nem sempre as melhores soluções estão entre as primeiras páginas ranqueadas, já que entre elas, existem aquelas irrelevantes para desenvolvedores. Para superar as limitações citadas acima, neste trabalho é proposto uma abordagem para minerar soluções relevantes para tarefas de programação a partir de resultados de serviços de busca. A abordagem funciona da seguinte forma: primeiro, é elaborada uma consulta relacionada à tarefa de programação, em seguida, esta é dada como entrada para um serviço de busca, na próxima etapa, as páginas retornadas são selecionadas e, finalmente, as páginas relevantes são selecionadas por meio de filtragem automática. Na avaliação, foram realizados cinco estudos de casos, com diferentes linguagens de programação. Primeiro, foram avaliadas as vinte primeiras páginas retornadas pelo serviço de busca do Google, para cada um dos estudos de casos, por meio dos seguintes critérios: aderência de conteúdo, exemplos práticos, simplicidade, grau de foco, entropia qualitativa e tamanho da solução. Depois, foram construídas quatro abordagens e realizadas comparações entre elas. Os resultados das avaliações das páginas retornadas pelo serviço de busca, para os cinco estudos de casos, mostram que em média, apenas 33% das páginas são relevantes para desenvolvedores que desejam aprender uma linguagem de programação e/ou uma API. O restante das páginas possui características indesejáveis, tais como: sem exemplos práticos, não relacionadas com a consulta realizada, com soluções extensas, sem foco no conteúdo pesquisado e soluções diferentes de outras retornadas. Os resultados das comparações entre as abordagens construídas, mostram que, a abordagem que utiliza ocorrências únicas de elementos técnicos nas páginas (como valor de atributo para o algoritmo de aglomeração), é a mais eficiente em remover páginas com características indesejáveis (em média, 83,1% das páginas indesejáveis foram removidas) e funciona com diferentes linguagens de programação e APIs, obtendo bons resultados.

Trilha: Trabalhos de Doutorado

Automated Extraction of Tumor Region of Interests from H&E-stained Oral Histological Whole Slide Images

Dali F. D. dos Santos (Universidade Federal de Uberlândia); Marcelo Zanchetta do Nascimento (Universidade Federal de Uberlândia); Bruno A. N. Travençolo (Universidade Federal de Uberlândia); Paulo Rogério de Faria (Universidade Federal de Uberlândia)

dalifreire@gmail.com; marcelo.zanchetta@gmail.com; travencolo@gmail.com;
paulorfaria1976@gmail.com

Resumo: Cancer can be defined as a group of non-communicable diseases (NCD) that can start almost anywhere in the human body. The disease is caused when cells begin to divide uncontrollably with potential to invade other parts of the body or spread to other organs and surrounding tissues. Projections from the National Cancer Institute (INCA) shows that in 2020, 686 thousand new cancer cases will be registered in Brazil and, from those cases, 15 thousand will be oral cavity-derived cancer. The histological slides evaluation, i.e., the study of tissue samples under the microscope, remains the most reliable way for diagnosing and grade oral cavity-derived cancer. This method of microscopic investigation, despite being referred as gold standard in cancer diagnosis, is plagued by inter- and intra-observer variability being a very time-consuming task that requires well-trained and experienced pathologists. Whole slide images (WSI) are multi-gigapixel images created from tissue sections that can contain many different types of cells and regions – such as keratin, lymphocytes, blood vessels, glands, muscle and tumor cells. WSI may also contain some derived artefacts from the image acquisition process, like tissue folding and blurring, that need to be ignored. Processing these huge images to discard non-relevant areas, keeping only clinically relevant regions, is a non-trivial problem. The vast majority of the existing approaches bypass this non-trivial problem by requiring a specialist to previously select, from the larger images, accurate small regions of interest (ROI) areas. Another frequent challenge is the scarcity of image datasets to study different types and tissue regions where oral cavity cancer can arise. A single H&E-stained histological WSI may contain thousands of nuclei and other cell structures that will be analyzed by pathologists at the diagnosis time. This large amount of information requires very skilled professionals and a great effort to label the relevant structures to create these image datasets. In this work, we proposes a novel methodology – using fully convolutional neural networks (FCN) – that can automatically analyze a WSI to localize and segment clinically relevant regions of oral cavity tumor to aid and enhance the pathologist decision making on diagnoses and prognoses of oral cavity-derived cancer. The proposed methodology uses hand-crafted features in a preprocessing step to remove background and scanning artefacts and a FCN to localize and segment 640x640 patch based tumor regions. To validate the proposed methodology, a new image dataset of H&E-stained tissues from the oral cavity region was created. The method was tested on this dataset achieving patch segmentation accuracy of 0.97 showing up robust with potential to help create a tool to increase efficiency and accuracy of pathologists detecting oral cavity-derived cancer in histological images.

Trilha: Trabalhos de Doutorado

Efficient Data Structures for Caching and Querying Large Temporal Graphs

Luiz Brito (Universidade Federal de Uberlândia); Marcelo Albertini (Universidade Federal de Uberlândia); Bruno Travençolo (Universidade Federal de Uberlândia); Arnaud Casteigts (Université de Bordeaux)

luiz.brito@ufu.br; albertini@ufu.br; travencolo@ufu.br; arnaud.casteigts@labri.fr

Resumo: Temporal graphs serve as a modeling tool to represent relationships that occurs over time. An important concept in temporal graphs is that of temporal paths among entities (times of consecutive relationships in path are strictly or non-strictly increasing). Based on that, we can think of high-level queries of utmost importance such as verifying whether an entity can reach another through a temporal path. Recent studies have applied temporal graphs and high-level queries to, respectively, abstract and solve problems in several contexts such as social networks and communication networks. However, when data grows unbounded, storing and querying large temporal graphs dynamically become difficult. Large in this context means that temporal graphs do not fit entirely into primary memory. On one hand, we need to continually persist data on high latency (and cheaper) storage and, on the other, we need to load these data into faster (and more expensive) memories using minimal space in order to process queries as fast as possible. Therefore, a good approach should consider all memory hierarchy in order to optimize high-level queries in temporal graphs. Particularly for primary memory, there is a growing area of study on compact data structures. These data structures consider space lower bounds based on Information Theory concepts, and provide useful queries with cost equal or close to standard data structures. In our scenario, these structures have great potential to be used as cache in faster memories, in order to reduce accesses of temporal graph data that are stored in slower storages. However, the majority of them are (1) static and therefore do not allow updates; or (2) are dynamic, but perform very poorly when compared to other static data structures (due to the RUM conjecture). In our work, we study techniques to make compact data structures a viable choice to be used as cache of temporal graph data that are stored in slower storages in order to answer high-level queries quickly. As our first contribution, we wrote a survey on compact data structures for temporal graphs (in review phase). As our second contribution, we studied an interesting problem in collaboration with the University of Bordeaux in France; develop a data structure that dynamically maintains the temporal reachability between entities of a temporal graph (in submitting phase). As next steps, we will investigate: (1) the feasibility of using compact data structures as cache for our dynamic data structure that answer temporal reachability queries; (2) the role of replacement policies for these new specialized cache structures; (3) whether we can use domain knowledge to improve performance of restricted classes of temporal graphs.

Trilha: Trabalhos de Doutorado