Ferramentas de Análise Estatística e Correlação para Sistemas de Recuperação da Informação

Khaick Oliveira Brito¹; Rodrigo Tripodi Calumby²

1. Bolsista PIBITI/CNPq, Graduando em Engenharia de Computação, UEFS, e-mail: khaickbrito@gmail.com

2. Orientador, Departamento de Ciências Exatas, UEFS, e-mail: rtcalumby@uefs.br

PALAVRAS-CHAVE: Recuperação de informação ; Eficácia; Análise estatística

INTRODUÇÃO

Atualmente, para efetuar o comparativo de sistemas de recuperação da informação (RI) usando medidas de eficácia, pesquisadores/desenvolvedores necessitam utilizar um conjunto de ferramentas para mensurar eficácia, realizar análise estatística, fazer análise de correlação para diferentes cenários e construir modelos visuais para apresentação dos resultados. Em especial, para garantir o rigor científico na análise comparativa de eficácia de diferentes métodos, é necessária a utilização de métodos para determinação da significância estatística considerando a execução dos métodos em múltiplas bases de dados e consultas. De modo geral, estas são tarefas trabalhosas dado que as ferramentas utilizadas em geral consomem e geram dados em diferentes formatos e possuem um conjunto limitado de funcionalidades. Neste trabalho, foi realizada a integração de funcionalidades de análise de significância estatística e correlação AnalyzIR, ferramenta que tem sido desenvolvida na UEFS e que visa disponibilizar ao usuário um ambiente integrado de análise de eficácia de sistema de recuperação de informação

MATERIAL E MÉTODOS

Novos recursos foram adicionados à ferramenta, como os testes de Mann-Whitney U, Correlação de Spearman, Correlação de Pearson, Coeficiente de Similaridade de Jaccard e que foram aprimorados, como *Wilcoxon's Signed Rank Test* e Teste t de Student. Os testes de significância estatística formulam duas hipóteses, nula e alternativa, que serão consideradas para concluir acerca das amostragens as quais o teste será aplicado, como por exemplo avaliar a existência de uma diferença significativa entre as médias dessas amostras.

Os testes de *Wilcoxon e Mann-Whitney U* podem ser aplicados para se comparar duas amostragens relacionadas ou emparelhadas visando verificar diferenças significativas, caso existam, entre os comportamentos dessas amostragens, podendo concluir se há chance de elas pertencerem à mesma população ou não. Ambos computam a diferença entre os valores emparelhados, a magnitude dessa diferença, em seguida ordenam essas diferenças e definem postos sinalizados. Por fim, aplica-se a fórmula estatística do teste em questão, encontrando um valor a ser comparado com o tabelado considerando o grau de liberdade e nível de significância escolhidos. Caso o valor encontrado seja superior ao tabelado, é indicado se rejeitar a hipótese nula de que as amostragens se assemelham ou suas médias são estatisticamente equivalentes.

O cálculo do valor t de *Student* se faz de maneira específica mediante a igualdade ou não da variância e da igualdade ou não entre os tamanhos das amostragens. que será comparado posteriormente aos valores encontrado na tabela de valores críticos de *t*, a partir de um grau de liberdade e um valor de confiança. O coeficiente de correlação de *Jaccard* é uma estatística usada para comparar conjuntos finitos e calcular um índice para a similaridade entre eles. O valor encontrado utilizando operações entre os conjuntos a serem comparados, fica inserido num intervalo [0, 1], em

que quanto maior o coeficiente mais alta a similaridade. O coeficiente de correlação de *Pearson* busca avaliar a existência de uma relação linear entre as variáveis apresentadas nas amostras. Através do cômputo de valores de variância e covariância, o teste de *Pearson* gera um valor p que indica o grau da correlação encontrado no intervalo $[-1, 1]$, em que valores próximos de 1 expressam uma correlação direta entre as variáveis analisadas e -1 para uma correlação inversa e 0 para uma não dependência entre elas. O teste de correlação de ordem de posto de *Spearman*, também como em *Pearson*, procura encontrar e apresentar o tipo de relação existente entre as variáveis contínuas ou ordinais apresentadas a partir de valores de covariância e desvio padrão.

Para efetuar o teste de *Spearman* devidamente na ferramenta, foi necessário desenvolver e aplicar um método de adaptação para *rankings* que não compartilhem de todos os elementos entre si, ou seja, que não tenha totalidade de intersecção nos elementos dos rankings, estimando o posto de um elemento não presente em um ranking, mas presente no outro, sendo mais distante o posto atribuído ao elemento ausente que se encontra mais próximo do topo do ranking ao qual ele pertence, de modo a permitir o cômputo do coeficiente. O método proposto se dá da seguinte forma:

- 1) Forma-se um conjunto com a união dos elementos dos *rankings* A e B, sendo A e B conjuntos com tamanhos iguais.
- 2) Criam-se dois novos vetores, A1 e B1, correspondentes à A e B.
Para cada elemento dessa união:
- 3) Verifica-se se ele existe em ambos os vetores originais A e B
- 4) Caso exista em algum deles, o valor da sua posição no seu vetor de origem. será guardado no vetor correspondente ao vetor original em que foi encontrado
- 5) Caso somente exista em um dos dois *rankings* originais, o posto atribuído a esse elemento no vetor correspondente ao *ranking* ao qual o elemento não faz parte será calculado pela Equação (1), onde T corresponde ao tamanho do ranking original e P é a posição ocupada pelo elemento no ranking original.

$$\text{Posto} = 2T - P \quad (1)$$

A Figura 1 demonstra a aplicação do método de adaptação em questão onde A e B são vetores de tamanhos iguais porém com alguns elementos iguais (A, B e C) e outros diferentes (D e E).

Figura 1: Exemplo de adaptação de *rankings*.

Tratando-se da aplicação desses testes em relação à ferramenta e à análise de sistemas de RI, os parâmetros necessários para se efetuar devidamente os testes são:

- 1) Testes de *Wilcoxon* e *Mann-Whitney U*: Escolha de dois sistemas a serem avaliados, um valor para o grau de confiança (e.g., 0.05), uma medida de avaliação, um valor de profundidade para os rankings e uma ou mais consultas.
- 2) Teste *t* de *Student*: Escolha de dois sistemas a serem avaliados, um valor para o grau de confiança (e.g., 0.05), uma medida de avaliação, um valor de profundidade para os rankings e uma ou mais consultas.
- 3) Teste de Spearman e coeficiente de *Jaccard*: Escolha de um ou mais sistemas, uma ou mais consultas e um valor de profundidade.
- 4) Teste de Pearson: Escolha de dois sistemas a serem avaliados, uma medida de avaliação, um valor de profundidade para os rankings e uma ou mais consultas.

RESULTADOS

A interface para os testes de *Wilcoxon* e *Mann Whitney U*, exibida na Figura 2A, apresenta os campos necessários para sua execução, bem como sua região de resultados. Existe a possibilidade de três versões para o teste de *Wilcoxon*, Apache Commons Math (ACM) [1], Java Statistical Classes (JSC) [2] e Mackenzie [3], e duas para *Mann Whitney U*, ACM e JSC, que podem ser escolhidas na interface apresentada.

As interfaces para os testes *t* de *Student* e correlação de *Pearson* se assemelham à apresentada na Figura 2A. A tela de configuração dos teste de *Jaccard* e *Spearman* pode ser vista na Figura 2B.

Figura 2: Telas de configuração **A** - Interface para os testes de *Wilcoxon* e *Mann Whitney U*; **B** - Interface para os testes de *Jaccard* e *Spearman*.

Além dos testes, foi disponibilizada também na ferramenta a possibilidade de visualização de resultados de alguns testes por meio de mapa de calor, apresentado na Figura 3, o qual permite edições em títulos, legendas e escala de cores. É possível realizar a exportação desse mapa de calor em formatos *png* e *jpg*.

Figura 3: Exemplo de mapa de calor gerado e exportado.

CONSIDERAÇÕES FINAIS

Ao término desse projeto de iniciação tecnológica, pôde-se contemplar os resultados obtidos satisfatoriamente, incluindo o aprimoramento da ferramenta AnalyzIR, que se deu pela incorporação do módulo estatístico desenvolvido nesse projeto, o que permitirá aos pesquisadores que fizerem uso da ferramenta efetuarem testes com bibliotecas conceituadas e suporte a múltiplos sistemas, o que permite um maior rigor para conclusões científicas, visualizações de resultados de forma mais intuitiva através de mapas de calor com legendas e cores customizáveis junto à exportação de resultados e tipos diferentes de arquivos, para se adequar às diferentes necessidades.

REFERÊNCIAS

- [1] COMMONS MATH: THE APACHE COMMONS MATHEMATICS LIBRARY. Disponível em: <<http://commons.apache.org/proper/commons-math/>> Acesso em: 15 ago. 2018.
- [2] CLASS WILCOXON SIGNED RANK. Scott MacKenzie. Disponível em: <<http://www.yorku.ca/mack/HCIbook/stats/WilcoxonSignedRank.html/>> Acesso em: 15 ago. 2018.
- [3] JAVA STATISTICAL CLASS. Disponível em: <www.jsc.nildram.co.uk> Acesso em: 15 out. 2017.