



Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Departamento de Ciências de Computação
SCC0284 – Sistemas de Recomendação

Trabalho 1

Professor: Dr. Marcelo G. Manzato (mmanzato@icmc.usp.br)
Estagiário PAE: Gustavo J. E. Ticona (gescobedo@usp.br)

Data da Entrega: 08/05/2018 até às 23:59:59.

1 Descrição

O objetivo deste projeto é implementar, testar e avaliar um conjunto de algoritmos de recomendação vistos em aula. Para isso, deverá ser utilizado o sistema Kaggle (<https://www.kaggle.com/c/scc0284-sistemas-de-recomendao>), a fim de que os resultados reportados na ferramenta possam ser utilizados no relatório deste projeto.

Nesta primeira fase, os seguintes algoritmos deverão obrigatoriamente ser desenvolvidos:

- Filtragem colaborativa baseada em itens;
- RF-Rec;
- Método probabilístico;
- Método baseline;
- Fatoração de matrizes (usando SVD da Álgebra Linear);
- Fatoração de matrizes (usando Gradiente Descendente Estocástico).

Para cada algoritmo, os grupos deverão submeter uma resposta no sistema Kaggle e coletar o RMSE gerado. Recomenda-se testar os algoritmos com diferentes valores dos parâmetros (e.g. número de vizinhos mais próximos, número de fatores latentes, taxa de aprendizado, regularização, etc.). Para que diferentes valores de parâmetros possam ser testados de maneira mais rápida, pode-se dividir a base de treinamento (obtida do Kaggle) em dois conjuntos, treino e validação, e usar o primeiro conjunto para treinar uma versão preliminar do algoritmo, e o segundo conjunto para mensurar o erro obtido (em termos de RMSE). Portanto, essa avaliação preliminar é executada localmente, sem envolver o sistema Kaggle. Uma vez determinada a melhor combinação dos parâmetros do algoritmo, o modelo poderá ser treinado novamente (usando toda a base disponível e os parâmetros definidos), a fim de gerar as predições que serão submetidas ao sistema Kaggle. O critério de divisão da base de treinamento deve ser justificado no relatório.

A entrega deste projeto consiste em dois pacotes:

1. **Código-fonte gerado no desenvolvimento dos algoritmos.** Os grupos são livres para escolher qualquer linguagem de programação desejada, porém, os seguintes aspectos serão avaliados neste quesito: documentação e organização interna (comentários, nomes de variáveis, clareza/-legibilidade); documentação e organização externa (passo-a-passo de como compilar e executar o programa, configurações necessárias do computador e compilador, requisitos mínimos do sistema); modularização do programa (separação lógica/física dos módulos, reutilização de código, etc.).

2. **Relatório de desempenho.** Os grupos deverão elaborar um relatório contendo o desempenho de cada algoritmo no sistema Kaggle, em termos de acurácia (RMSE) e tempo de execução (separar as fases de treinamento e predição). Neste relatório, deverá ser relatado também como foi feita a seleção de parâmetros de cada algoritmo, bem como os resultados obtidos no caso de implementação de outros algoritmos não listados nesta especificação, ou variações nos algoritmos especificados que obtiveram sucesso. Adicionalmente, deverão ser discutidos os fenômenos observados no sistema (e.g. limitações, possíveis melhorias, etc.).

2 Instruções Complementares

- São permitidos grupos formados por até dois integrantes;
- Os algoritmos desenvolvidos deverão ser implementados pelo grupo, ou seja, a utilização de código proveniente da Internet ou cópia entre grupos não serão toleradas;
- Os resultados do relatório deverão corresponder ao desempenho mensurado pelo sistema Kaggle. Portanto, diferentes submissões deverão ser submetidas no sistema, e para cada uma, o RMSE resultante deverá ser anotado pelo grupo;
- O código-fonte e o relatório deverão ser compactados em formato zip, e o arquivo resultante deverá ser submetido no sistema Tidia-Ae (<http://ae4.tidia-ae.usp.br/>);
- Haverá um desconto de 1 ponto na nota do trabalho por dia de atraso a partir da data da entrega.

Bom Trabalho!