

Estudo Preliminar da Construção de um Modelo de Recomendação de Algoritmos de Seleção de Atributos utilizando Meta-Aprendizado

Antonio Rafael Sabino Parmezan, Huei Diana Lee, Wu Feng Chung
Laboratório de Bioinformática – LABI, UNIOESTE, Foz do Iguaçu, PR

Objetivos

Realizar uma investigação preliminar sobre a viabilidade de utilizar Meta-Aprendizado (MA) [1], baseado em características intrínsecas aos Conjuntos de Dados (CD) e dos algoritmos de Seleção de Atributos (SA) analisados experimentalmente em trabalhos anteriores [2], como auxílio para a construção de um modelo de recomendação de algoritmos de SA.

Métodos/Procedimentos

O processo de construção do Modelo de Recomendação (MR) foi organizado em quatro etapas. Na etapa 1, foram selecionados 30 CD naturais obtidos do repositório de dados UCI¹. Após, esses CD foram submetidos à remoção de exemplos com valores desconhecidos. Na etapa 2, cada CD passou, inicialmente, pela extração das medidas dispostas na Tabela 1.

Tabela 1 – Medidas de caracterização de CD

Medidas simples	Medidas estatísticas
Número de atributos	Erro da classe majoritária
Número de atributos qualitativos	Assimetria média e desvio padrão
Número de atributos quantitativos	Curtose média e desvio padrão
Número de exemplos	Correlação média e desvio padrão
Número de classes	Dimensão fractal do CD

Em seguida, foi realizada a SA usando as medidas de consistência e correlação (algoritmos *Consistency-basedFilter* (CBF) e *Correlation-basedFeatureSelection* (CFS)). A partir dos subconjuntos de atributos selecionados, foram induzidos modelos de árvores de decisão utilizando o algoritmo J48 e, baseado no Desempenho Preditivo (DP), identificou-se o algoritmo de SA mais adequado para cada CD. A medida DP foi estimada por meio de Validação Cruzada (VC) com 10 partições. Dessa etapa, obteve-se um atributo alvo nominal que pode assumir os valores “CBF” ou “CFS”. Na etapa 3, os meta-exemplos foram constituídos por Meta-Atributos de Entrada (MAE) e Meta-Atributos Alvo (MAA) obtidos na etapa 2. Na etapa 4, ocorreu o mapeamento entre os MAE e os MAA usando, novamente, o J48 para a construção do meta-modelo que

contempla o conhecimento embutido nos meta-exemplos e que, posteriormente, será utilizado para auxiliar na recomendação de algoritmos de SA. O DP do MR foi estimado também por meio de VC com 10 partições. Para a aplicação do método proposto foi construído um sistema computacional que provê suporte à exploração e remoção de valores faltantes, à construção da meta-base e à recomendação de algoritmos.

Resultados

O MR construído apresentou cinco regras de classificação. Observou-se que do total de 13 atributos, apenas quatro constituíram o meta-modelo. Considerando como critério de importância a frequência com que aparecem os MAE nas regras induzidas, os atributos mais importantes foram: assimetria média dos atributos e o respectivo desvio padrão. É importante ressaltar que a média e o desvio padrão do DP do modelo foram 56,33 e 27,09, respectivamente, ou seja, obteve-se um erro de 43,66, próximo ao da classe majoritária (43,33).

Conclusões

Neste trabalho, procurou-se verificar a viabilidade de encontrar propriedades dos CD que pudessem indicar a adequação de determinados algoritmos de SA. Apesar do MR ter apresentado baixo DP, possibilitou constatar a importância do emprego de medidas estatísticas para a caracterização de CD. Trabalhos futuros incluem a avaliação de outros métodos para construção de sugestões e algoritmos para indução de meta-modelos, bem como a análise de outras medidas de caracterização.

Referências Bibliográficas

- [1] Brazdil PB; Giraud-Carrier C; Soares C; Vilalta R. *Metalearning: applications to data mining*. Springer, 2009.
- [2] Parmezan, ARS; Wu, FC; Lee, HD. Estudo de medidas de importância e algoritmos para SA para mineração de dados. XX EAIC, 2011.

¹Frank A and Asuncion A. UCI Machine Learning Repository. Irvine: University of California, School of Information and Computer Science, 2010. Disponível em: <<http://archive.ics.uci.edu/ml>>. Acesso em: 20 jun. 2012.