

Reduzindo discriminação em classificadores

Hilton Pintor Bezerra Leite

Victor Miranda de Melo

Uma das aplicações da mineração de dados é a tomada de decisão. Para isso podem ser construídos Classificadores, que através de aprendizagem com dados de treinamento resultam num modelo de Árvore de Decisão utilizado para classificar novos dados. Porém quando treinados com bases de dados que foram construídas num contexto discriminativo, i.e., quando elementos de um grupo são privados de benefícios e oportunidades somente por pertencerem ao grupo, o modelo resultante será enviesado, perpetuando a discriminação, e podendo com o avanço da sociedade vir a perder sua acurácia.

Estudos provaram que apenas remover atributos considerados sensíveis dos dados de treinamento com o intuito de evitar a discriminação nos classificadores não é suficiente, pois os mesmos tendem a manter esse comportamento através da percepção do relacionamento com outros atributos não removidos. Foram então propostas técnicas baseadas em pré-processamento dos dados anterior ao treinamento, e técnicas que são sensíveis à discriminação presente nos dados durante o processo de criação do modelo, essas últimas se mostrando mais eficientes.

Nas Árvores de Decisão, os nós dividem o conjunto de dados em partições distintas de acordo com critérios relacionados aos valores dos atributos iterativamente, até que numa folha possa ser feita a classificação do dado de entrada com um certo grau de confiança. Tradicionalmente, o critério de partição é dado pela sua contribuição para o aumento da acurácia da previsão, porém nos modelos sensíveis à discriminação também é levado em conta sua influência no grau de discriminação causado pela partição.

Outra abordagem proposta é a Remarcação das Folhas, onde dado uma Árvore de Decisão são trocadas as classes de algumas folhas selecionadas, de forma com que a discriminação seja diminuída com o mínimo de perda de acurácia. Geralmente diminuir a discriminação ocasiona em uma diminuição da precisão, por isso a Remarcação das Folhas faz uso de um algoritmo guloso baseado no KNAPSACK para se aproximar de uma remarcação ótima.

Por fim foi observado através de experimentação comparativa com os métodos do estado da arte, que abordagens que combinam Remarcação de Folhas com critérios de de partição sensíveis à discriminação, conseguem diminuir consideravelmente a taxa de discriminação, enquanto mantém a acurácia alta. Além disso, quando usados conjuntos de dados enviesados para treinamento, em

conjunto com dados de teste com menos discriminação, essas abordagens conseguem também maior precisão.

Em outro estudo, foram criados três novas técnicas: *Modified Naive Bayes*, *2 Naive Bayes Models* e *Latent Variable Model*. Esses alteram o classificador Naive Bayes visando eliminar a discriminação encontrada no conjunto de dados.

O Modified Naive Bayes modifica a distribuição de Bayes de forma que aumenta a probabilidade para casos sensíveis discriminados e diminui probabilidade de casos sensíveis favorecidos. Removendo classes até que não haja mais label discriminatória. Esse procedimento, elimina a discriminação, porém continua com a influência de um atributo discriminatório em outros atributos.

O 2 Naive Bayes Models trabalha evitando as dependências entre atributos, removendo a correlação entre atributos sensíveis e não sensíveis. O algoritmo evita remover os atributos sensíveis para não haver perda muito grande na acurácia, removendo apenas o fato que eles podem ser usados para decidir.

Por último, Latent Variable Model é uma abordagem não observada, onde é criada uma variável Latente que é independente dos atributos sensíveis e os rótulos são livres de discriminação. Além disso, ele assume que cada classe tem chances iguais de serem discriminadas, sendo independentes dos atributos não sensíveis, consequentemente independente da probabilidade de ser rotulado positivamente.

Como resultado dos testes, 2 Naive Bayes obteve a melhor performance, atingindo a melhor acurácia com nenhuma discriminação e tem a menor dependência dos atributos sensíveis.