Print

Relational Logistic Regression

A third example of a relational or network classifier is relational logistic regression. This was first introduced by Lu and Getoor in 2003. It combines both local and network (or link) attributes. Remember that local attributes describe a customer's behavior such as age, income, recency, frequency, monetary characteristics, and so on. Various types of network attributes are then also included, such as the mode-link, which considers the most frequently occurring class among the neighbors; the count-link, which looks at the frequency of the classes of the neighbors; and the binary link, which adds binary indicators for class presence. Both local and network attributes can then be combined into a single logistic regression model. Note that the idea of adding network features to the data is commonly referred to as featurization or propositionalization. Relational logistic regression gave good performance in the study by Macskassy and Provost. It is also very commonly applied in the industry.

Here is an example of the featurization process for relational logistic regression. Customer Bart is characterized by two local characteristics, age and income. A first network feature is the mode-link. In this example, the mode-link is non-fraudulent because three of the five neighbors are non-fraudsters. Two frequency features are added: one for the non-fraudsters, which equals 3, and one for the fraudsters, which equals 2. Finally, two binary network features are added: one for the non-fraudsters, which equals 1, and one for the fraudsters, which also equals 1. Notice that the network features are obviously correlated, so it's important to do input selection to select the most predictive features. For example, you can use stepwise logistic regression.

Close