Automatic Resource Access Control Project Proposal

Xinghuang Xu

**Project Title:**

Automatic Resource Access Control

**Project Idea:**

Employee in companies need to access resources like read/manipulate various applications or web portals to fulfill their job responsibilities. Usually there is a knowledgeable supervisor who would manually verify the request and grant the access in order to overcome access obstacles. This access discovery/recovery process could waste a nontrivial amount of time and money especially in a company that employees move frequently.

In big corporations, there is considerable amount of data regarding an employees’ role within the organization and the resources to which they have access. Given the data related to current employees and their provisioned access, can we build models to automatically grant/recovery access privileges as employees enter/leave roles within a company? The goal of the model is to minimize the human factor in granting/revoking employee access.

**Software:**

This goal of the project is not to implement any specific machine learning algorithms but to quickly construct models using pre implemented machine learning algorithms as building blocks. The final tool would take in a training data set with N entries. Each entry contains the resource category, information regarding the role and the final action. After the model building, the tool would be able to predict the action, grant or revoke, given the resource requested and the information about the role.

**Dataset:**

It’s extremely hard to obtain information regarding a company’s employees’ role information and their provisioned access because those are considered top secret of the company and can provide severe security threads once known to the public. Fortunately, there is a dataset provided by amazon at the Kaggle data science competition site that we can use to test our model. After signing the agreement, I was able to download the dataset and it has more then 30,000 entries each with 9 columns, the first column contains the final action, the second column tells about the resource category and the other 7 columns are features regarding the role.

**Papers to read:**

Cannot find any papers solving this specific problem. As a supervised learning problem, most of the supervised learning techniques like decision tree, logistic regression, naïve bayes and etc can be used. The hard part is to figure out which algorithm performs better and why.