A Supervised Approach to Predicting Company Funding Success with Factual and Topic Features

F. Qaiser

## Introduction

Raising financing is one of the important aspects of corporate strategy for almost any startup. Financing provides the economic fuel needed to develop prototypes, hire employees, produce marketing materials etc. Understanding the underlying factors that affect funding is therefore useful information for guiding entrepreneurs along their journey.

## The problem

Previous work in this area has been limited in part because of its reliance on mostly numerical features such as accounting, financial and market variables in building models. As startups are typically not obligated to record and make public their financial statements, such information is available only in limited quantities.

## The solution

Luckily, today we have access to a vast body of textual information from a variety of sources (such as social media sites, news portals, \_\_\_\_ etc.), which often discuss the latest in tech trends. There is a strong possibility that this information could also be mined for predicting startup success/failure at raising funding.

To explore the value of this information, we will use the profiles and news articles for companies and people on CrunchBase, the largest public database for the tech world, which anybody can edit. CrunchBase provides access to this data up to the end of 2013 freely. Specifically, we will attempt to explore a few financial variables alongside topic features via topic modeling techniques within a machine learning framework.

## Deliverables

The final deliverable will be the code I used, a paper with my results and maybe a slide deck.

## Approach

I’m less interested in interpretation and more interested in accuracy at this stage. For the approach, this means I want to mostly focus on using machine learning techniques rather than regression modelling.

Also, currently I intend to predict whether a company will receive funding or not.

My suspicion however is that the data on CrunchBase is likely to be biased towards companies that do receive funding and is unlikely to be a representative sample of the population. (There are of course techniques I could use to address this of course but I’m not super interested in those kinds of challenges).

In the event that this proves true, we may have to revise our project to predicting how much funding a company will receive.

Or even switch to predicting whether a company wil exit (i.e. acquisition or IPO).