A Supervised Approach to Predicting Company Exits with Factual and Topic Features

F. Qaiser

## Introduction

Ultimately, every startup journey ends in some kind of an exit. Broadly speaking, this can be an IPO, an acquisition or a closure.

An exit is one of the important aspects of corporate strategy for almost any startup founder as they offer the chance to cash in years of hard work in the form of their equity holdings. Understanding the underlying factors that affect exit outcomes is therefore useful information for guiding entrepreneurs along their journey.

Investors too are also likely to be interested in the same information given that venture capital as an asset class has historically generated poor returns. For them, the real value would stem from being able to identify accurately (and early on) startups that show signs of making it big.

This topic is also likely to be of interest to the wider population at large. In recent years, big exits that have turned hundreds of early employees of a few startups into overnight millionaires.

## 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.

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.

## Things that may change

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.

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).