A DAG of Latent Topic Models for Multimodal Relational Data

Yi Wang, Guan Wang, Ramanuja Simha

July 15, 2014

We argue in this article that we can make full use of multimodal relational data to address difficulties of high-dimensionality, sparsity and cold-start. We propose to build a deep nerual network by chaining up latent topic/factor models, e.g., latent Dirichlet allocation or restricted Boltamann machine, learned from relational data.

An example of three relational data, company–founder, company–follows and follows–post, is illustrated in Figure 1. Suppose that we want to predict companies' performance (measured by value evaluations or finance reports) using features derived from the data.

A straightforward approach is that for each company, use its founders and follows as features. However, this is well-known of suffering from data sparseity. Existing multimodal modeling techniques would decompose company–founder and company–follower matrices using latent topic models and use t_1 and t_2, latent topics learned from these matrices respectives, as features. This solves the sparsity problem somehow, but still suffers from the difficulty of cold-start.

Consider a cold-start problem that we want to predict the performance of a new company c, which is not in the training data illustrated in Figure 1. Suppose that the two founders, f_1 and f_2, of c exist in our training data, so we can use them as features, or use them to derive latent topics in type t_1 as features. However, suppose that the two followers, l_1 and l_2, of c are not in the training data. This means that there is no way to use them or latent topics of type t_2 as features. However, we noticed that l_1 like posts $\mathbf{o}_1$ and that l_2 like posts $\mathbf{o}_2$. Can we use this to address this cold-start problem?

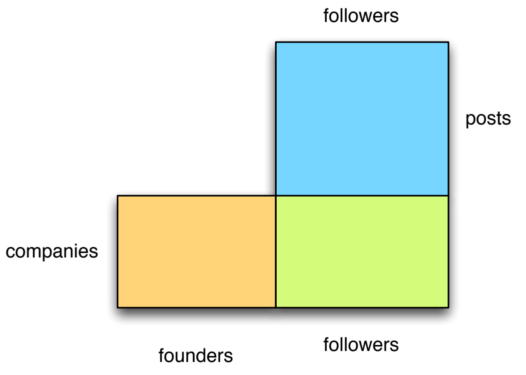

Figure 1: Relational Data

We think we can. Our idea is to learn a latent topic model from each matrix in Figure 1. As illustrated in Figure 2, learning from the company-founder matrix produces count matrix $C(company, t_1)$, the number of times that a latent topic t_1 is assigned to a company, and $C(t_1, founder)$, the number of times that a founder founded a company for a reason (latent topic). It is noticable that count matrices can be used with Bayesian inference to understand new data. This can help us addressing the cold-start problem.

Consider the aforementioned cold-start problem. Given $C(t_3, post)$ as illustrated in Figure 2, we can infer $P(t_3|\mathbf{o}_1)$ and $P(t_3|\mathbf{o}_2)$. This allows us to find some followers who are in the training data, or some latent topics of type t_2, as features in our prediction. Denote followers by f, we have

$$P(t_2|\mathbf{o}_1, \mathbf{o}_2) = \sum_f \sum_{t_3} P(t_2|f)P(f|t_3) \left[\sum_j P(t_3|\mathbf{o}_j) \right] \tag{1}$$

Since summations like $\sum_{t_3}$ and $\sum_f$ can be intuitively interpreted as activations in neural networks, we can visualize above equation using a deep neural network, as illustrated in Figure **??**.

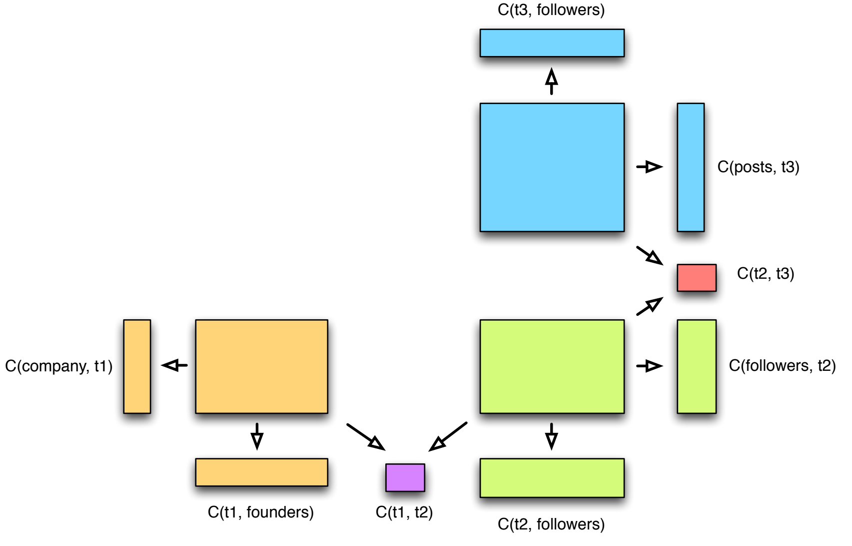

Figure 2: Decomposition