

尊敬的吴老师，您好

我想先用一个例子，简要介绍一下训练过程，以及论文中一处不合理的地方。

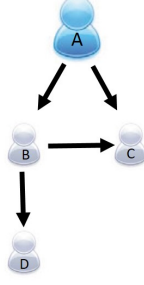


图 1: 微博用户网络示例

如图1所示(黑色箭头表示“关注”关系), 假设对于一条微博 m , 它的转发链为

$$\{A, B, C, D\}$$

据此, 我们可以得到每个用户的上下文链, 如表1所示。

用户	$i = 1$	$i = 2$	$i = 3$	$i = 4$
A	$\{\}$	$\{\}$	$\{\}$	$\{\}$
B	$\{A\}$	$\{A\}$	$\{A\}$	$\{A\}$
C	$\{A\}$	$\{A, B\}$	$\{A, B\}$	$\{A, C\}$
D	$\{\}$	$\{B\}$	$\{B\}$	$\{B\}$

表 1: 用户上下文链 $D_{v,i}^m$

此时的状态向量为

$$z_A^m = [1, 1, 1, 1, 1]^T$$

$$z_B^m = [0, 1, 1, 1, 1]^T$$

$$z_C^m = [0, 0, 1, 1, 1]^T$$

$$z_D^m = [0, 0, 0, 1, 1]^T$$

接下来需要求解 $P(z_v | \delta)$, 我们以 $P(z_B | \delta)$ 作为示例。

由于 B 并非微博 m 的始发者, 因此

$$p(z_{B,0}^m = 0) = 1 - p(z_{B,0}^m = 1) = 1$$

根据 $z_B^m = [0, 1, 1, 1, 1]^T$ 和 $\delta(A, B) = 1$ (根据图1), 可以得到

$$p(z_{B,1}^m = 1 | z_{B,0}^m = 0, D_{B,1}^m, \delta) = 1 - \exp(-\lambda \delta(A, B) \sum_{u \in D_{B,1}^m} I_u^T S_v) = 1 - \exp(-\lambda I_A^T S_v)$$

$$p(z_{B,2}^m = 1 | z_{B,1}^m = 1, D_{B,2}^m, \delta) = 1$$

$$p(z_{B,3}^m = 1 | z_{B,2}^m = 1, D_{B,3}^m, \delta) = 1$$

$$p(z_{B,4}^m = 1 | z_{B,3}^m = 1, D_{B,4}^m, \delta) = 1$$

而 $P(z_B^m | \delta)$ 即为它们相乘的结果，即

$$P(z_B^m | \delta) = 1 - \exp(-\lambda I_A^T S_v)$$

用户A、C、D状态向量 z_v^m 的概率的求法与上面类似，最终我们可以得到 $L(C)$ 。当然，我们在训练过程中并不需要计算 $L(C)$ ，只需要计算 $\frac{\partial L}{\partial I_u}$ 和 $\frac{\partial L}{\partial S_v}$ ，然后利用投影梯度法更新 I 和 S 。不断重复上述过程，直到论文的公式(6)达到最小值。

$$L(C) = \prod_{m=1}^{|C|} \prod_{v \in V} P(z_v^m | \delta)$$

假如我们还有很多微博在图1所示的网络中传播，我们不难发现用户的上下文链有大量的重叠。为了避免冗余计算，所以论文中提出了公式(7)，将所有消息的上下文链按用户分组。

总结一下训练过程：

- 1) 构建diffusion network，以求解 $\delta(u, v)$
- 2) 对每一条消息，求解上下文链 $D_{v,i}^m$ ，并对用户分组，得到 $D_{v,i}$
- 3) 对每一条消息，求解状态向量 $z_{v,i}$
- 4) 按照论文中Algorithm 1计算 I, S

原论文中一个不合理的地方

论文中的公式(4)是这样的

$$p(z_{v,i}^m = 1 | z_{v,i-1}^m = 0, D_{v,i}^m, \delta) = 1 - \exp(-\lambda \delta(a_i^m, v) \sum_{u \in D_{v,i}^m} I_u^T S_v)$$

其中的 $\delta(a_i^m, v)$ 的 a_i^m 表示消息 m 的转发链中的第 i 个用户，这在真实网络中明显不合理！

举个例子，对于图1，某时刻用户A作为始发者，转发了一条信息 m ，B、C作为粉丝，受到A的影响，也转发了此消息，且B要比C先转发。此时得到的转发链为 $\{A, B, C\}$ 。根据公式(4)计算 $p(z_{C,2}^m = 1 | z_{v,1}^m = 0, D_{v,2}^m, \delta)$ 时，公式(4)中的 $\delta(a_i^m, C)$ 为 $\delta(B, C)$ 。但我们知道C之所以转发信息，是受A的影响，而非B！这样可能成本应属于A的影响力部分的转移到了B身上。再者，假如B到C之间没有连边，那么 $\delta(B, C) = 0$ ，最终导致公式(6)计算的 $L(C)$ 趋于无穷大。

这可能是作者的笔误，也有可能是我对文章的理解不到位。

学生王超民，2016年5月10日