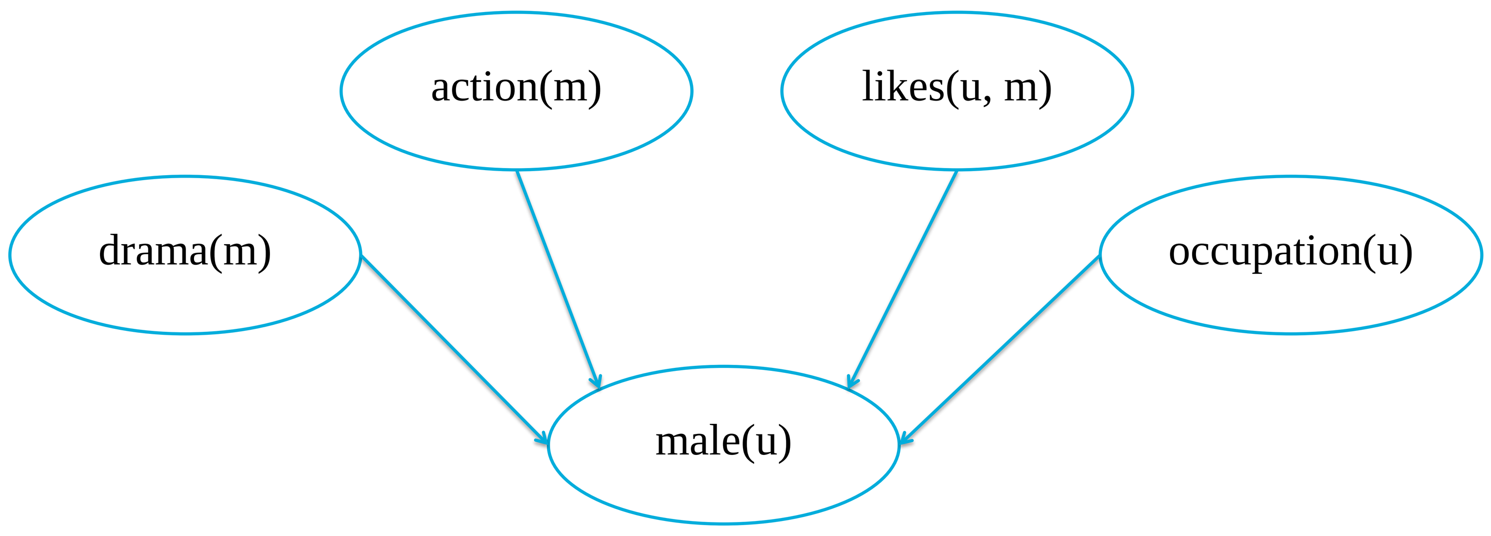
A Learning Algorithm for Relational Logistic Regression

Preliminary Results

Bahare Fatemi, Seyed Mehran Kazemi & David Poole

Relational Logistic Regression (RLR)

- Relational Logistic Regression (RLR) defines probability distributions using weighted formulas (WFs).



- Example WFs for defining the conditional probability distribution (CPD) of male(u):

$$true: w_0$$
$$likes(u,m) \wedge action(m): w_1$$
$$likes(u,m) \wedge drama(m): w_2$$
$$occupation(u) = doctor: w_3$$

- Sample data:

	$likes(u,m)$			
	m_1	m_2	m_3	m_4
u_1	☑	☑	☑	☒
u_2	☒	☒	☑	☑

u	$occupation(u)$
u_1	homemaker
u_2	doctor

m	$action(m)$
m_1	☒
m_2	☒
m_3	☑
m_4	☑

m	$drama(m)$
m_1	☑
m_2	☑
m_3	☒
m_4	☒

u	$male(u)$
u_1	☒
u_2	☑

- Sufficient statistics for $u = u_1, u_2$:

$$true = 1$$
$$\#m: likes(u_1,m) \wedge action(m) = 1$$
$$\#m: likes(u_1,m) \wedge drama(m) = 2$$
$$(occupation(u_1) = doctor) = 0$$

$$true = 1$$
$$\#m: likes(u_2,m) \wedge action(m) = 2$$
$$\#m: likes(u_2,m) \wedge drama(m) = 0$$
$$(occupation(u_2) = doctor) = 1$$

- The model implies P(male(u)) is:

$$P(male(u_1)) = sig(w_0 * 1 + w_1 * 1 + w_2 * 2 + w_3 * 0)$$
$$P(male(u_2)) = sig(w_0 * 1 + w_1 * 2 + w_2 * 0 + w_3 * 1)$$

- In general:

$$P(male(u)) = sig(w_0 * 1 + w_1 * (\#m: likes(u,m) \wedge action(m)) + w_2 * (\#m: likes(u,m) \wedge drama(m)) + w_3 * (occupation(u) = doctor))$$

Parameter Learning

- Convert the data into a form that can be used by traditional ML learners for each CPD.
- Example of the converted data for CPD of $male(u)$

u	$true$	$m: likes(u,m) \wedge action(m)$	$m: likes(u,m) \wedge drama(m)$	$occupation(u) = doctor$	$male(u)$
u_1	1	1	2	0	$false$
u_2	1	2	0	1	$true$

- Using logistic regression, we can learn the weights of these formulas.

Structure Learning

- Hierarchical assumption:
  - A WF is defined to be *useless* if
    - ❑ An L-1 regularized learner learned a zero weight for it, or
    - ❑ A subset of it is useless
- Search Strategy
  - We start with WFs with at most one unary atom like:

$$true: w_0$$
$$likes(u,m) \wedge action(m): w_1$$
$$likes(u,m) \wedge \neg action(m): w_2$$
$$likes(u,m) \wedge drama(m): w_3$$
$$likes(u,m) \wedge \neg drama(m): w_4$$
$$occupation(u) = homemaker: w_5$$
$$occupation(u) = doctor: w_6$$

  - Consider L1-regularized logistic regression learns the weights as:

$$w_0, w_1, w_3 \neq 0, and\ w_2, w_4, w_5, w_6 = 0$$

Based on the hierarchical assumption non-zero weights indicate useful WFs and zero weights indicate useless WFs.

$$true: w_0 \neq 0 \Rightarrow useful$$
$$likes(u,m) \wedge action(m): w_1 \neq 0 \Rightarrow useful$$
$$likes(u,m) \wedge \neg action(m): w_2 = 0 \Rightarrow useless$$
$$likes(u,m) \wedge drama(m): w_3 \neq 0 \Rightarrow useful$$
$$likes(u,m) \wedge \neg drama(m): w_4 = 0 \Rightarrow useless$$
$$occupation(u) = homemaker: w_5 = 0 \Rightarrow useless$$
$$occupation(u) = doctor: w_6 = 0 \Rightarrow useless$$

  - Conjoining useful WFs gives the feature below:

$$likes(u,m) \wedge drama(m) \wedge action(m): w_7$$

  - Updated list of WFs:

$$true: w_0$$
$$likes(u,m) \wedge action(m): w_1$$
$$likes(u,m) \wedge drama(m): w_3$$
$$likes(u,m) \wedge drama(m) \wedge action(m): w_7$$

  - Hierarchical assumption forbids the below WF:

$$likes(u,m) \wedge action(m) \wedge occupation(u) = doctor$$

Hidden Features

- Each object may contain useful information that has not been observed
  - Example: in predicting gender of people, some movies may only be appealing to males and some only to females
- Learning Hidden Features
  - We add continues unary atoms such as H(m)
  - Random initial values
  - Learn their values using stochastic gradient descent using L1-regularization

Results

- Dataset: Movielens data-set with the modification made by Schulte and Khosravi
- Data contains 940 users and 1682 movies
- Predicting age and gender of users based on their attributes, the movies they have rated (ignoring the actual rating) and movie attributes
- The learning algorithms are:
  - Baseline: predicting the mean of the data
  - LR is the standard logistic regression
  - RLR-Base: RLR without use of hidden Features
  - RLR-H: RLR with considering one hidden feature.
- Accuracy: the percentage of correctly classified instances (higher accuracies indicate better performance)
- Average conditional log likelihood (ACLL) is computed as follows:

$$ACLL = \frac{1}{m}\sum_{i=1}^{m} \ln(P(male(u_i)|data, model))$$ (higher accuracies indicate better performance)

		Learning Algorithm				
		Baseline	LR	RDN-Boost	RLR-Base	RLR-H
Gender	ACLL	-0.6020	-0.5694	-0.5947	-0.5368	-0.5046
	Accuracy	71.0638%	71.383%	70.6383%	73.8298%	77.3404%
Age	ACLL	-0.6733	-0.5242	-0.5299	-0.5166	-0.5090
	Accuracy	60.1064%	76.0638%	76.4893%	77.1277%	77.0212%