ZooBP: Belief Propagation for Heterogenous Networks

Dhivya Eswaran*
CMU
deswaran@cs.cmu.edu

Stephan Guennemann
TUM
guennemann@in.tum.de

Christos Faloutsos
CMU
christos@cs.cmu.edu

Disha Makhija
Flipkart
disha.makhiji@flipkart.com

Mohit Kumar
Flipkart
k.mohit@flipkart.com
e-commerce fraud detection
Inferencing in social networks

**Groups**

**Users**

**Pages**

Eswaran, Guennemann, Faloutsos, Makhija & Kumar
Heterogenous graph labeling

**GIVEN**

- a heterogenous graph with multiple node types & edge types
- labels/classes for a few nodes
- compatibility matrix per edge type* (network effects)

**FIND**

- labels/classes of all nodes
Compatibility matrices for modeling network effects

assume it is given!
Classification by propagation

**INITIALIZE:**
- Labeled nodes: “known” values
- Unlabeled nodes: “default/random” values

**PROPAGATE:**
- Update each node’s value based on the values of its neighbors.

**CONVERGENCE:**
- If no value changes, terminate.

Q1. What are values here?
Q2. How are they updated?
Belief propagation

Values: beliefs (probability vectors)

Update: 2 stages

(i) Send messages based on belief

(ii) Update belief based on messages
Belief update rules

MESSAGES FROM BELIEFS

\[ m_{vu}(i) \leftarrow \sum_j H(i, j)b_v(j)/m_{uv}(j) \]

BELIEFS FROM MESSAGES

\[ b_u(i) \leftarrow e_u(i) \prod_{v \in \mathcal{N}(u)} m_{vu}(i) \]

Eswaran, Guennemann, Faloutsos, Makhija & Kumar
Wishlist

- BP applies to heterogeneous graphs too but has no convergence guarantees.

**THEORY**
- T1. Closed form solution
- T2. Convergence guarantees

**PRACTICE**
- P1. Good approximation
- P2. Fast & scalable
What conditions allow approximation?

**A1** Constant margin compatibility matrix

Row sum = 3 (const)
Col sum = 2 (const)

```
+ 0.2  x
```

Interaction strength
Residual compatibility matrix

**A2** Low interaction strength

Eswaran, Guennemann, Faloutsos, Makhija & Kumar
Linearizing approximations

- **Logarithm approximation**
  \[ \log(1 + x) \approx x \]

- **Division approximation**
  \[ \frac{a + x}{1 + y} \approx a + x - ay \]
Linearized BP update rules

**MESSAGES FROM BELIEFS**

\[
m_{uv}^{(t)}(i) \leftarrow \frac{\epsilon_t}{k_s} \sum_j H_t(i, j) \left( k_s b_v(j) - m_{uv}^{(t)}(i) \right)
\]

- **Res. message I send**
- **Res. compatibility**
- **My current res. belief**
- **Echo cancellation**

**BELIEFS FROM MESSAGES**

\[
b_u(i) \leftarrow e_u(i) + \frac{1}{k_s} \sum_{v \in \mathcal{N}_u} \sum_{t \in T_{uv}} m_{vu}^{(t)}(i)
\]

- **Your updated res. belief**
- **Your prior res. belief**
- **Res. messages you receive**
✓ Problem
✓ Background
✓ Approximations

ZOOBP

EXPERIMENTS
A layer of abstraction: personas

“Every class label for every node is a persona”

What we want to solve for: Persona residual belief vector

# personas: $N = 2 \times n_u + 3 \times n_p$

Users’ residual beliefs

Pages’ residual beliefs

Eswaran, Guennemann, Faloutsos, Makhija & Kumar
ZooBP in a single slide

\[ b = e + (P - Q)b \]

- final persona
- residual belief
- prior persona
- residual belief
- persona influence
- echo cancellation

function of connectivity structure and compatibility matrices
Persona influence

\[ P = \begin{bmatrix} P_{11} & \ldots & P_{1S} \\ \vdots & \ddots & \vdots \\ P_{S1} & \ldots & P_{SS} \end{bmatrix} \]

\[ P_{ss'} = \sum_{t \in T_{ss'}} \frac{e_t}{k_s} (H_t \otimes A_t) \]
Echo cancellation

\[ Q = \begin{bmatrix} Q_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Q_s \end{bmatrix} \]

\[ Q_s = \sum_{s' \in S} \sum_{t \in T_{ss'}} \frac{\epsilon_t^2}{k_s k_{s'}} (H_t H_t^T \otimes D_{st}) \]
ZooBP in a single slide (again)

\[ b = e + (P - Q)b \]

ITERATIVE UPDATE: \[ b \leftarrow e + (P - Q)b \]
Wish list

**THEORY**

**T1.** Closed form solution

**T2.** Convergence guarantees

**PRACTICE**

**P1.** Good approximation

**P2.** Fast & scalable

\[ b = e + (P - Q)b \]
Closed form solution

\[
b = e + (P - Q)b
\]

\[
b = (I + Q - P)^{-1}e
\]

STRICTLY DIAGONALLY DOMINANT
Wish list

**THEORY**

- **T1.** Closed form solution
- **T2.** Convergence guarantees

**PRACTICE**

- **P1.** Good approximation
- **P2.** Fast & scalable
ZooBP convergence guarantees

\[ b = (I + Q - P)^{-1}e \]

\[ \rho(P - Q) < 1 \]

CLOSED FORM

CONVERGENCE CONDITION

SPECTRAL NORM (LARGEST EIGEN VALUE)

GRAPH STRUCTURE

COMPATIBILITY

Eswaran, Guennemann, Faloutsos, Makhija & Kumar
Wish list

**THEORY**
- T1. Closed form solution
- T2. Convergence guarantees

**PRACTICE**
- P1. Good approximation
- P2. Fast & scalable
PROBLEM

BACKGROUND

APPROXIMATIONS

ZOOBP

EXPERIMENTS
Datasets

**DBLP (authors, papers)**
- 2 node types, 1 edge type
- $k=2$ to 7; areas/conferences

**Flipkart (users, products, sellers)**
- 3 node types, 5 edge types
- $k=2$ for each node type
Wish list

**THEORY**

- **T1.** Closed form solution
- **T2.** Convergence guarantees

**PRACTICE**

- **P1.** Good approximation
- **P2.** Fast & scalable
ZooBP matches BP’s belief scores

Flipkart (users, products)
ZooBP detects e-commerce fraudsters

Flipkart

~100% @ 250
~70% @ 500
Wish list

THEORY

- T1. Closed form solution
- T2. Convergence guarantees

PRACTICE

- P1. Good approximation
- P2. Fast & scalable

Eswaran, Guennemann, Faloutsos, Makhija & Kumar
ZooBP is faster than BP
Wish list

**THEORY**
- T1. Closed form solution
- T2. Convergence guarantees

**PRACTICE**
- P1. Good approximation
- P2. Fast & scalable
Summary

- **A1** Constant margin compatibility
- **A2** Low interaction strength

\[ b = (I + Q - P)^{-1}e \]

- **T1.** Closed form solution
- **T2.** Convergence guarantees
- **P1.** Good approximation
- **P2.** Fast & scalable

Questions? deswaran@cs.cmu.edu