Activity-Edge Centric Multi-label Classification for Mining Heterogeneous Information Networks

Yang Zhou and Ling Liu

Distributed Data Intensive Systems Lab (DiSL)
College of Computing
Georgia Institute of Technology
Our Approach

- **Labeling w.r.t. multiple activity networks**
  - Separate different types of activity graphs from the heterogeneous information networks
  - Exploit the correlations among the set of class labels within each activity graph and across multiple activity graphs

- **Label dependency w.r.t. multiple activity networks**
  - Tackling label dependency should take into account of important information hidden in the label space; and design a learning algorithm to model inter-dependencies among multiple labels

- **Vertex-centric classification vs. edge-centric classification**
  - Combine both vertex-centric classification and edge-centric classification to boost the effectiveness and efficiency

- **Structure affinity vs. label vicinity**
  - Incorporate both structure affinity and label vicinity into a unified classifier to speed up the classification convergence

- **AEClass is an iterative learning algorithm**
  - Dynamically refine the classification result by continuously learning the contribution of multiple edge classification schemes, structure affinity and label vicinity in the unified classifier.
Various Information Networks Involved in AE-Class

- Collaboration graph $CG$
  - Model relationships between people
  - Example: Facebook, LinkedIn, DBLP

- Activity graph $AG_i$ ($1 \leq i \leq N$)
  - Model relationships between activities
  - Activities can be any type of objects or things
  - Example: books linked by topic categories, smart phones linked by manufactures

- Collaboration multigraph $MG$
  - An activity-edge augmented multigraph of $CG$
  - Model parallel relationships between vertices in $CG$
  - For each edge in $CG$, create a set of parallel edges between the pair of vertices. Each set of unlabeled edges in $CG$ has up to $K$ labeled parallel edges in $MG$ and each parallel edge corresponds to one activity category labeled by $c_p$ in each $AG_i$
An Illustrating Example from DBLP

Coauthor Graph

Ming-Syan Chen  Wei Fan
Kun-Lung Wu  Charu C. Aggarwal

Philip S. Yu

17 32
4
16
17

Conference Graph

VLDB  CDM
SIGMOD  KDD
ICDE  SDM

0.00393  0.00408
0.00392  0.00464
0.000977  0.000977
0.00947
0.00947
0.000937  0.000897
0.000967
0.000905
0.00407  0.00442
0.00408  0.00460

Coauthor Multigraph

Ming-Syan Chen  Wei Fan
Kun-Lung Wu  Charu C. Aggarwal

Philip S. Yu

1 1
16 16
4
12
3
14

2 - DB Conference  2 - DM Conference
Problem Definition

• Given a collaboration graph $CG$ and multiple activity graphs $AG_i (i=1,\ldots,m)$, we want to classify the vertices in $CG$ into the $K$ classes by
  – Utilizing the relationships between vertices in the collaboration graph
  – Utilizing the relationships between nodes in each activity graph
  – Constructing a collaboration multigraph $MG$

• Achieve a good balance between vertex-centric classification and edge-centric classification

• Achieve a good balance between structure affinity and label vicinity

• Classification algorithm should be fast and scalable to the size of collaboration graph and the number and size of activity graphs
Algorithm Framework

- Input: One collaboration graph and multiple activity graphs
- Step 1: Partition Each Activity Graph into Clusters
- Step 2: Activity-based Edge Classification
- Step 3: Collaboration Multigraph Construction
- Step 4: Activity-edge Centric Vertex Classification
- Step 5: Improvement by Edge Label Dependency
- Step 6: Refinement by Vertex Label Vicinity
- The above process iterates until the classification objective converges
Step 1: Partition Each Activity Graph into Clusters

Conference Graph

DM Conference Cluster

DB Conference Cluster

Term Graph

DM Term Cluster

DB Term Cluster
Edge Splitting

• For each activity graph $AG_i$ and the collaboration graph $CG$, construct an activity-edge augmented collaboration multigraph $CG_i$ by examining each edge and the pair of connected entities in $CG$ and splitting each edge into a set of parallel edges based on each activity in $AG_i$ that this pair of entities have in common.
Step 2: Activity-based Edge Classification

- The class-membership probability of edge \((v_p, v_q)\) in \(CG_i\) belonging to class \(c_j\) based on \(AG_i\)

\[
P(L_{pq} = c_j|AG_i) = \frac{1}{W(p,q)} \sum_{m=1}^{N_{AG_i}} W_i^m(p,q)P(L_m = c_j|AG_i)
\]

- Reduce \(CG_i\) to \(\overline{CG}_i\) with classified edges by grouping at most \(N_{AG_i}\) parallel edges between any pair of vertices into at most \(K\) parallel edges

\[
\overline{W}_j(p,q) = W(p,q)P(L_{pq} = c_j|AG_i)
\]
Step 3: Collaboration Multigraph Construction

- Combine $N$ edge classification schemes of $CG$, i.e., $CG_i$ into a unified $MG$ with different weighting factors $\omega_1, \ldots, \omega_N$

  $$W_{j}^{(t)}(p, q) = \sum_{i=1}^{N} \omega_i^{(t)} W_j^{i}(p, q)$$

  $$= \sum_{i=1}^{N} \omega_i^{(t)} \sum_{m=1}^{NAG_i} W_m(p, q) P(L_m = c_j|AG_i), \quad 1 \leq j \leq K$$

  subject to $\sum_{i=1}^{N} \omega_i^{(t)} = 1$, $\omega_i^{(t)} \geq 0$, $i = 1, \ldots, N$

- Transition probability on $MG$

  $$T_{j}^{(t)}(p, q) = \begin{cases} 
  \frac{W_{j}^{(t)}(p, q)}{\sum_{r=1}^{NCG} \sum_{m=1}^{K} W_m^{(t)}(p, r)}, & \text{for } p > l, \\
  1, & \text{for } p = q \leq l, \\
  0, & \text{otherwise}.
\end{cases}$$
Step 4: Activity-edge Centric Vertex Classification

- Initial unified classification kernel (lack of label vicinity)
  \[ K_j^{(1)} = T_j^{(1)} \]

- Block matrix form
  \[
  K_j^{(1)} = \begin{bmatrix}
  K_{jll}^{(1)} & K_{jlu}^{(1)} \\
  K_{jul}^{(1)} & K_{juu}^{(1)}
  \end{bmatrix} = \begin{bmatrix}
  I & 0 \\
  K_{jul}^{(1)} & K_{juu}^{(1)}
  \end{bmatrix}
  \]

- Iterative classification
  \[
  X_j^{(t)} = K_j^{(t)} X_j^{(t-1)}
  \]
  \[
  X_j^{(t-1)} = [X_{jl}^{(t-1)}, X_{ju}^{(t-1)}],
  X_j^{(t)} = K_{jul}^{(t)} X_{jl}^{(t-1)} + K_{juu}^{(t)} X_{ju}^{(t-1)}
  \]

- Class-membership matrix
  \[
  X^{(t)} = [X_1^{(t)}, X_2^{(t)}, \cdots, X_K^{(t)}] = 
  \begin{bmatrix}
  X_{1l}^{(t)} & X_{2l}^{(t)} & \cdots & X_{Kl}^{(t)} \\
  X_{1u}^{(t)} & X_{2u}^{(t)} & \cdots & X_{Ku}^{(t)}
  \end{bmatrix}
  \]
Step 5: Improvement by Edge Label Dependency

- Edge label dependency by activity category similarity between \( c_p \) and \( c_q \) wrt \( AG_i \)

\[
S_i(c_p, c_q) = \begin{cases} 
\sum_{u_m \in U_{ip}, u_n \in U_{iq}} \frac{S_i(u_m, u_n)}{|U_{ip}| \times |U_{iq}|}, & p \neq q, \\
1, & p = q.
\end{cases}
\]

- Adjust class-membership vector

\[
Y^{(t)}_{ju} = \sum_{m=1}^{K} \sum_{i=1}^{N} \omega_i^{(t)} S_i(c_j, c_m) X^{(t)}_{mu}, \quad 1 \leq j \leq K
\]

- Adjusted class-membership matrix

\[
Y^{(t)} = [Y_1^{(t)}, Y_2^{(t)}, \ldots, Y_K^{(t)}] = \begin{bmatrix}
X_{1l}^{(t)} & X_{2l}^{(t)} & \cdots & X_{Kl}^{(t)} \\
Y_{1u}^{(t)} & Y_{2u}^{(t)} & \cdots & Y_{Ku}^{(t)}
\end{bmatrix}
\]
Step 6: Refinement by Vertex Label Vicinity

- label vicinity \( (T^{(t+1)}_j Y^{(t)}) (T^{(t+1)}_j Y^{(t)})^T \) quantitatively measures the similar extent between vertices and their neighbors based on the current labeling.

- Example
  - both Philip S. Yu and his many coauthors, and Jiawei Han and his many coauthors have higher probabilities on DM, lower probabilities on DB, and zero probabilities on OS, then the label vicinity between two experts should be very large.

- Adjust unified classification kernel by incorporating structure affinity and label vicinity

\[
K^{(t+1)}_j = \alpha^{(t+1)} T^{(t+1)}_j + \beta^{(t+1)} (T^{(t+1)}_j Y^{(t)}) (T^{(t+1)}_j Y^{(t)})^T \\
\text{subject to } \alpha^{(t+1)} + \beta^{(t+1)} = 1, \alpha^{(t+1)}, \beta^{(t+1)} \geq 0
\]
Why AEClass Outperforms Existing Representative Multi-label Classification Models

• Accuracy improvement
  – Existing classifiers summarize all neighbors and all associated links to execute the inference
  – AEClass only aggregate those neighbors and associated links with the same label as the current objective class to execute the inference

• Efficiency improvement
  – No matter which class the current objective is, existing classifiers need to check each neighbor of a vertex and summarize the labels of all neighbors
  – AEClass only needs to consider those links with the same label $c_j$ as the current objective and associated neighbors and further stops label propagation to the circle of those irrelevant neighbors (without link with label $c_j$) in the future iterations
Classification Objective Function

- Graph classification objective function

\[
Macro-F1 \approx \frac{1}{K} \sum_{j=1}^{K} \frac{2 \sum_{i=l+1}^{N_{CG}} Y(i,j)y_i^j}{\sum_{i=l+1}^{N_{CG}} Y(i,j) + \sum_{i=l+1}^{N_{CG}} \theta y_i^j}
\]

- F1 score is the harmonic mean of precision and recall. Macro-F1 score is the unweighted mean of F1 score on classes

\[
Macro-F1 \approx \frac{\sum_{i=1}^{m} a_i(\alpha)^{bi}(\beta)^{ci} \prod_{j=1}^{N} (\omega_j)^{d_{ij}}}{\sum_{i=1}^{n} o_i(\alpha)^{pi}(\beta)^{qi} \prod_{j=1}^{N} (\omega_j)^{r_{ij}}},
\]

\[a_i, b_i, c_i, d_{ij}, o_i, p_i, q_i, r_{ij} \geq 0, b_i, c_i, d_{ij}, p_i, q_i, r_{ij} \in \mathbb{Z}\]

- The goal of multi-label classification of multigraph is to maximize the Macro-F1 score

\[
\max_{\alpha, \beta, \omega_j} \frac{\sum_{i=1}^{m} a_i(\alpha)^{bi}(\beta)^{ci} \prod_{j=1}^{N} (\omega_j)^{d_{ij}}}{\sum_{i=1}^{n} o_i(\alpha)^{pi}(\beta)^{qi} \prod_{j=1}^{N} (\omega_j)^{r_{ij}}} = \max_{\alpha, \beta, \omega_1, \ldots, \omega_N} \frac{f(\alpha, \beta, \omega_1, \ldots, \omega_N)}{g(\alpha, \beta, \omega_1, \ldots, \omega_N)}
\]

subject to \(\alpha + \beta = 1, \alpha, \beta \geq 0, \sum_{j=1}^{N} \omega_j = 1, \omega_j \geq 0, j = 1, \ldots, N\)
Parameter-based Optimization

• Transform a sophisticated nonlinear fractional programming problem of multiple weights $\alpha$, $\beta$, $\omega_1$, $\omega_2, \ldots, \omega_N$ into a straightforward nonlinear parametric programming problem of single variable $\gamma$

$$F(\gamma) = \max_{\alpha, \beta, \omega_1, \ldots, \omega_N} f(\alpha, \beta, \omega_1, \ldots, \omega_N) - \gamma g(\alpha, \beta, \omega_1, \ldots, \omega_N)$$

subject to $\alpha + \beta = 1$, $\alpha, \beta \geq 0$, $\sum_{i=1}^{N} \omega_i = 1$, $\omega_i \geq 0$, $i = 1, \ldots, N$

• Properties of $F(\gamma)$
  – Continuous
  – Convex
  – Monotonic decreasing
  – $F(\gamma) = 0$ has a unique solution
Activity-Edge Centric Multi-label Classification

1: Initialize $\alpha^{(2)}=\beta^{(2)}=\frac{1}{2}$, $\omega^{(1)}_1=\cdots=\omega^{(1)}_N=\frac{1}{N}$ and $\gamma^{(1)}=0$;
2: Partition each of $N$ kinds of activities into $K$ clusters simultaneously;
3: Calculate the category similarity on each $AG_i$;
4: Execute the edge classification on $CG$ based on each $AG_i$;
5: Construct the collaboration multigraph $MG$;
6: Compute $W_j^{(1)}$, $T_j^{(1)}$ and $K_j^{(1)}$ for each $c_j$;
7: Repeat the following steps until $F(\gamma^{(t)})$ converges to 0
8: Calculate the class-membership matrices $X^{(t)}$ and $Y^{(t)}$;
9: Compute the Macro-F1 score;
10: Solve the NPPP of $F(\gamma^{(t)})$;
11: Update $\alpha^{(t+1)}$, $\beta^{(t+1)}$, $\omega^{(t+1)}_1$, $\cdots$, $\omega^{(t+1)}_N$;
12: Refine $\gamma^{(t+1)}=f(\alpha^{(t+1)}, \beta^{(t+1)}, \omega^{(t+1)}_1, \cdots, \omega^{(t+1)}_N)/g(\alpha^{(t+1)}, \beta^{(t+1)}, \omega^{(t+1)}_1, \cdots, \omega^{(t+1)}_N)$;
13: Update $W_j^{(t+1)}$, $T_j^{(t+1)}$ and $K_j^{(t+1)}$;
Experimental Evaluation

• Datasets
  – **DBLP bibliography data**
    • 100,000 authors, 712,834 links, and two activity graphs: Conference and Term
    • 24 research areas: AI, AIGO, ARC, BIO, CV, DB, DIST, DM, EDU, GRP, HCI, IR, ML, MUL, NLP, NW, OS, PL, RT, SC, SE, SEC, SIM, WWW
  – **Last.fm music-oriented online social network**
    • 50,000 users, 496,611 links, and the two activity networks: Artist and Track
    • 21 music genres: acoustic, ambient, blues, classical, country, electronic, emo, folk, hardcore, hip hop, indie, jazz, latin, metal, pop, pop punk, punk, reggae, rnb, rock, soul
  – **IMDb internet movie database**
    • 10,000 actors, 270,227 links, and a Movie activity graph
    • 22 movie genres: Action, Adventure, Animation, Biography, Comedy, Crime, Documentary, Drama, Family, Fantasy, Film-Noir, History, Horror, Music, Musical, Mystery, Romance, Sci-Fi, Sport, Thriller, War, Western

• Baseline methods
  – LBC [Lu and Getoor, ICML'03]: one-hop structure information
  – wvRN [Mackassy and Provost, MRDM'03]: multiple-hop structure information
  – EdgeCluster [Tang and Liu, CIKM'09]: one-hop structure information
  – SCRN [Wang and Sukthankar, KDD'13]: multiple-hop structure information + EdgeCluster
Classification Quality Evaluation

Classification Quality on DBLP

(a) Macro-F1  
(b) Micro-F1  
(c) Hamming Loss

Classification Quality on Last.fm

(a) Macro-F1  
(b) Micro-F1  
(c) Hamming Loss
Clustering Efficiency

(a) DBLP
(b) Last.fm
(c) IMDb

Classification Efficiency

(a) Macro-F1
(b) Micro-F1
(c) Hamming Loss

Classification Quality on IMDb
Classification Convergence

(a) DBLP

(b) Last.fm

Classification Convergence
Weight Update

(a) DBLP

Weight Update

(b) Last.fm
# Case Study

<table>
<thead>
<tr>
<th>Author/Class</th>
<th>DB</th>
<th>DM</th>
<th>ML</th>
<th>IR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peter L. Bartlett</td>
<td>0.019</td>
<td>0.039</td>
<td>0.938</td>
<td>0.004</td>
</tr>
<tr>
<td>Elisa Bertino</td>
<td>0.738</td>
<td>0.054</td>
<td>0.028</td>
<td>0.180</td>
</tr>
<tr>
<td>Andrei Z. Broder</td>
<td>0.037</td>
<td>0.097</td>
<td>0.015</td>
<td>0.851</td>
</tr>
<tr>
<td>Michael J. Carey</td>
<td>0.965</td>
<td>0.029</td>
<td>0.006</td>
<td>0.023</td>
</tr>
<tr>
<td>W. Bruce Croft</td>
<td>0.054</td>
<td>0.007</td>
<td>0.037</td>
<td>0.902</td>
</tr>
<tr>
<td>David J. DeWitt</td>
<td>0.912</td>
<td>0.057</td>
<td>0.004</td>
<td>0.027</td>
</tr>
<tr>
<td>Inderjit S. Dhillon</td>
<td>0.030</td>
<td>0.409</td>
<td>0.457</td>
<td>0.104</td>
</tr>
<tr>
<td>Christos Faloutsos</td>
<td>0.321</td>
<td>0.500</td>
<td>0.031</td>
<td>0.147</td>
</tr>
<tr>
<td>Jiawei Han</td>
<td>0.391</td>
<td>0.463</td>
<td>0.045</td>
<td>0.100</td>
</tr>
<tr>
<td>H. V. Jagadish</td>
<td>0.850</td>
<td>0.056</td>
<td>0.009</td>
<td>0.048</td>
</tr>
<tr>
<td>Michael I. Jordan</td>
<td>0.007</td>
<td>0.062</td>
<td>0.917</td>
<td>0.014</td>
</tr>
<tr>
<td>Daphne Koller</td>
<td>0.026</td>
<td>0.045</td>
<td>0.915</td>
<td>0.013</td>
</tr>
<tr>
<td>Vipin Kumar</td>
<td>0.120</td>
<td>0.622</td>
<td>0.199</td>
<td>0.059</td>
</tr>
<tr>
<td>Bing Liu</td>
<td>0.086</td>
<td>0.427</td>
<td>0.266</td>
<td>0.220</td>
</tr>
<tr>
<td>Hector Garcia-Molina</td>
<td>0.788</td>
<td>0.010</td>
<td>0.016</td>
<td>0.186</td>
</tr>
<tr>
<td>C. J. van Rijsbergen</td>
<td>0.003</td>
<td>0.051</td>
<td>0.024</td>
<td>0.922</td>
</tr>
<tr>
<td>Michael Stonebraker</td>
<td>0.946</td>
<td>0.013</td>
<td>0.007</td>
<td>0.034</td>
</tr>
<tr>
<td>Jeffrey D. Ullman</td>
<td>0.824</td>
<td>0.065</td>
<td>0.064</td>
<td>0.047</td>
</tr>
<tr>
<td>Philip S. Yu</td>
<td>0.342</td>
<td>0.496</td>
<td>0.044</td>
<td>0.118</td>
</tr>
<tr>
<td>Mohammed J. Zaki</td>
<td>0.148</td>
<td>0.672</td>
<td>0.057</td>
<td>0.123</td>
</tr>
</tbody>
</table>

Class-membership Probabilities of Authors Based on Conference and Keyword Partitions from DBLP
Conclusions

• Integrate the primary social network and multiple associated activity networks into a unified multigraph with edge classification

• Combine both structure affinity and label vicinity based on multiple activity networks into a unified classifier

• An iterative learning algorithm is proposed dynamically refine the classification result by continuously adjusting the weights on different activity-based edge classification schemes from multiple activity graphs, while constantly learning the contributions of the structure affinity and the label vicinity in the unified classifier
Questions?

Thank you!

yzhou@gatech.edu
lingliu@cc.gatech.edu