Neural Message Passing for Multi-Label Classification

Jack Lanchantin, Arshdeep Sekhon, Yanjun Qi
University of Virginia, Department of Computer Science
Crisis on Wall Street as Lehman Totters, Merrill Seeks Buyer, AIG Hunts for Cash

U.S. Opts to Avoid Lehman Rescue, Stirring a Momentous Weekend for American Finance; Traders Brace for a Chaotic Monday

BY CARRICE MOLLENKAMP, SUSANNE CRAIG AND SERENA NG

The American financial system was shaken to its core on Sunday. Lehman Brothers Holdings Inc. faced the prospect of liquidation and Merrill Lynch & Co. was close to a deal to sell itself to Bank of America Corp. The U.S. government, which bailed out Fannie Mae and Freddie Mac a week ago and orchestrated the sale of Bear Stearns Cos. to J.P. Morgan & Co. earlier this year, drew a line in the sand with Lehman. It refused to provide a financial backstop to potential buyers.

Without such support, Barclays PLC and Bank of America, the two most interested buyers, walked away. On Sunday night, Bank of America was close to striking a deal to buy Merrill Lynch for about $44 billion, or $22 a share. Lehman was working on a possible bankruptcy filing. As worries spread across Wall Street that Lehman wouldn't survive, brokerage firms, hedge funds and other traders moved to disentangle themselves from trades with Lehman. When hopes of a potential sale dimmed, a quiet Sunday on Wall Street turned into a mad rush. Executives and traders hurried to their offices or worked their phones to unwind outstanding contracts with Lehman and to gauge their overall exposure. A sense of foreboding gripped the Street as top executives feared collateral damage from a Lehman liquidation. Attention turned to Merrill Lynch, which boasts the largest force of retail brokers, and to American International Group Inc., the insurer giant. Both firms have seen their stocks get hammered, and their managements spent the weekend trying to come up with plans to reassure the markets.

"Monday will be a day of reckoning for the financial markets," said Carlos Mendez, senior managing director of J.C. Capital, a boutique investment firm in New York. On Sunday, he said, "it was like a fire alarm went off and people ran in all directions." AIG executives spent the weekend trying to raise cash, either from asset sales or a capital infusion from private-equity firms, or both. AIG executives were meeting with regulators to see if they could transfer capital from some of its subsidiaries to the holding company. Merrill, whose retail brokerage force is the biggest in the country and is known as the "thundering herd," quietly engaged in discussions with Bank of America, whose retail bank branches stretch coast to coast. Wall Street executives said the Federal Reserve encouraged them in orchestrating the sale, figuring that it was "better to save the relatively healthy patient instead of the dying one," said a lawyer involved in the discussions. "We are in uncharted waters here," said a top executive of a big bank. "If Merrill can pull off a deal this weekend, that would certainly help." The U.S. dollar, which had strengthened in the past few weeks, fell against all four of its major rivals on Sunday — the euro, the Swiss franc, the U.K. pound and the Japanese yen. Some executives involved in the Lehman discussions held out hope that an 11th-hour reprieve would materialize. Under one scenario aimed at limiting the ripple effects of Lehman's demise, a group of about 15 banks were in discussions Sunday to pool about $100 billion, which would be used to buy assets of the beleaguered securities firm, according to one person familiar with the situation. Details were being finalized Sunday night. One possibility is that the Federal Reserve will support the move by opening its wholesale-borrowing window and relaxing collateral requirements for borrowers.

Ultimatum

By Paulson Sparked Frenzied End

After more than a century and a half as an investment house, Lehman Brothers Holdings Inc. tottered on the brink Sunday night. But its latest troubles were set in motion days earlier.

In midweek, Treasury Secretary Henry M. Paulson Jr. threatened a "huge" recapitalization that would have carried a "tremendous" cost, according to a bank official familiar with the discussions.

As Lehman's woes persisted, the U.S. government stepped in and initiated talks to sell the investment bank. Lehman and AIG were both in the running for a deal, but the negotiations broke down after AIG put the deal on hold. Lehman was ultimately acquired by Barclays PLC for $5.7 billion.

By Deborah Solomon, Dennis K. Berman, Suzanne Craig and Carrice Mollenkamp

Please turn to page A18
Crisis on Wall Street, U.S. Opt s to Avoid Lehman Rescue, Stirring a Momentous Weekend for American Finance; Traders Brace for a Chaotic Monday

BY CARRICE MOLLENKAMP, SUSANNE CRAIG AND SERENA NG

The American financial system was shaken to its core on Sunday, as Lehman Brothers Holdings Inc. faced the prospect of liquidation and Merrill Lynch & Co. was close to a deal to sell itself to Bank of America Corp. The U.S. government, which bailed out Fannie Mae and Freddie Mac a week ago and orchestrated the sale of Bear Stearns Cos. to J.P. Morgan & Co. earlier this year, drew a line in the sand with Lehman. It refused to provide a financial backstop to potential buyers. Without such support, Barclays PLC and Bank of America, the two most interested buyers, walked away. On Sunday night, Bank of America was close to striking a deal to buy Merrill Lynch for about $44 billion, or $29 a share. Lehman was working on a possible bankruptcy filing.

As worries spread across Wall Street that Lehman wouldn’t survive, brokerage firms, hedge funds and other traders moved to disentangle themselves from trades with Lehman. When hopes of a potential sale dimmed, a quiet Sunday on Wall Street turned into a mad rush. Executives and traders hurried to their offices or worked their phones to unwind outstanding contracts with Lehman and to gauge their overall exposure.

A sense of foreboding gripped the Street as top executives feared collateral damage from a Lehman liquidation. Attention turned to Merrill Lynch, which boasts the largest force of retail brokers, and to American International Group Inc., the insurer giant. Both firms have seen their stocks get hammered, and their management spent the weekend trying to come up with plans to reassure the markets.

Monday will be a day of reckoning for the financial markets, said Carlos Mendez, senior managing director of ICP Capital, a boutique investment firm in New York. On Sunday, he said, “it was like a fire alarm went off and people ran in all directions.”

AIG executives spent the weekend trying to raise cash, either from asset sales or a capital infusion from private-equity firms, or both. AIG executives were meeting with regulators to see if they could transfer capital from some of its subsidiaries to the holding company.

Merrill, whose retail brokerage business is a key component in the country and is known as the “thundering herd,” quietly engaged in discussions with Bank of America, whose retail bank branches stretch coast to coast. Wall Street executives said the Federal Reserve was working in orchestrating the sale, figuring that it was “better to save the relatively healthy patient instead of the dying one,” said a lawyer involved in the discussions.

We are in uncharted waters here,” said a top executive of a big bank. “If Merrill can pull off this deal this weekend, that would certainly help.”

The U.S. dollar, which had strengthened in the past few weeks, fell against all four of its major rivals on Sunday—the euro, the Swiss franc, the U.K. pound and the Japanese yen.

Some executives involved in the Lehman discussions held out hope that an 11th-hour reprieve would materialize. Under one scenario, aimed at limiting the ripple effects of Lehman’s demise, a group of about 15 banks were in discussions Sunday to pool about $100 billion, which would be used to buy assets of the bailed-out securities firm, according to one person familiar with the situation. Talks were being finalized on Sunday night. One possibility is that the Federal Reserve will support the move by opening its wholesale-borrowing window and relaxing collateral requirements for borrowers.

Lehman, a 158-year-old firm Please turn to page A18
Crisis on Wall Street, Lehman Totters, Merrill Seeks Buyer, AIG Hunts for Cash

The American financial system was shaken to its core on Sunday when Lehman Brothers Holdings Inc. faced the prospect of liquidation and Merrill Lynch & Co. was close to a deal to sell itself to Bank of America Corp. The U.S. government, which bailed out Fannie Mae and Freddie Mac a week ago and orchestrated the sale of Bear Stearns Cos. to J.P. Morgan & Co. earlier this year, drew a line in the sand with Lehman. It refused to provide a financial backstop to potential buyers. Without such support, Barclays PLC and Bank of America, the two most interested buyers, walked away. On Sunday, Lehman was working on a possible bankruptcy filing.

As worries spread across Wall Street that Lehman wouldn't survive, brokerage firms, hedge funds and other traders moved to distance themselves from trades with Lehman. When hopes of a sale dimmed, a quick Wall Street turnaround rush. Executives hurried to their offices to place their phones to unloading promising contracts with the easy money firms, or both. AIG executives were meeting with regulators to see if they could transfer capital from some of its subsidiaries to the holding company.

Merrill, whose retail brokerage force is involved in the mortgage business in the country and “Thundering the noise of the financial markets,” said Carlos Mendez, senior managing director. “The macro picture is a little bit more deflationary than expected.”

The U.S. dollar, which had strengthened in the past few weeks, fell against all four of its major rivals on Sunday—the euro, the Swiss franc, the yen, pound and the Japanese yen.

Ultimate End

By Paulson Sparked Frenzied End

After more than a century and a half as an investment house, Lehman Brothers Holdings Inc. tumbled on the by now familiar script. It's latest troubles were set in motion days earlier. In midweek, Treasury Secretary Henry Paulson announced the government would support the move by opening its wholesale-borrowing window and relax collateral requirements for borrowers.

Lehman, a 158-year-old firm, Please turn to page A19.
Multi Label Classification (MLC)

- MLC is the task of assigning a set of target labels for a given sample.
- Given input $x$, predict the set of labels $\{y_1, y_2, ..., y_L\}$, $y_i \in \{0,1\}$.
Binary Relevance Classifiers

\[ p(y|x) \approx \prod_{i=1}^{L} p(y_i|x) \]
Probabilistic Chain Classifiers
Read et. al. 2009, Wang et. al 2016, Nam et. al. 2017

\[ p(y|x) = \prod_{i=1}^{L} p(y_i|y_{1:i-1}, x) \]
Deep Structured Prediction Classifiers
Belanger & McCallum 2016, Gygli et. al. 2017

\[
y = \arg \min_{\tilde{y}} E(\tilde{y}; F(x)) \quad \tilde{y} \in \{0,1\}^L
\]
Embedding Models

Bhatia et. al. 2015, Wu et. al. 2017

\[ \text{sim}(x, y_1), \text{sim}(x, y_2), \ldots \]
probabilistic chain classifiers

embedding and energy function classifiers
probabilistic chain classifiers

This Work

embedding and energy function classifiers

Interpretable

Fast
Background: Message Passing Neural Networks
Message Passing Neural Networks (MPNNs)

Scarselli et. al. 2009, Gilmer et. al. 2017

\[ G = (V,E) \]

\[ v_i^t \in \mathbb{R}^d \]
Message Passing Neural Networks (MPNNs)
Scarselli et. al. 2009, Gilmer et. al. 2017

- MPNNs can learn useful node representations for classifying nodes by encoding local graph structures and node attributes
- **Main idea:** pass messages between pairs of nodes and update them

\[
G = (V,E)
\]

\[
\mathbf{v}_i^t \in \mathbb{R}^d
\]
Message Passing Neural Networks (MPNNs)
Scarselli et. al. 2009, Gilmer et. al. 2017
Message Passing Neural Networks (MPNNs)

Scarselli et. al. 2009, Gilmer et. al. 2017

message function $F_m$

$$m_i^t = \sum_{j \in N(i)} F_m(v_i^t, v_j^t),$$

node update function $F_u$

$$v_i^{t+1} = F_u(m_i^t)$$
Joint representations of nodes and edges are modelled using message passing rather than explicit probabilistic formulations, allowing for efficient inference.
MPNNs for Multi Label Classification
Neural Message Passing for MLC
Neural Message Passing for MLC

\[
\begin{align*}
    \text{input} & : x \\
    \text{city} & : u_2^t \\
    \text{road} & : u_3^t \\
    \text{car} & : u_1^t \\
    \text{castles} & : u_4^t \\
    \text{mountains} & : u_5^t
\end{align*}
\]
Neural Message Passing for MLC
This Paper: Label Message Passing (LaMP) Networks

- **Main Idea:** Labels are represented as nodes in a label-interaction graph.
- Given input embedding $x$, the goal of LaMP is to model the conditional dependencies between label embeddings $\{u^t_{1:L}\}$ using Message Passing Neural Networks.
- LaMP uses MPNN modules to update label embeddings at each step $t$ in two parts.
Part A: Input-to-Label Message Passing

The first step passes messages from input $\mathbf{x}$ to each label embedding $\mathbf{u}_i^t$ to produce the intermediate embedding state $\mathbf{u}_i^{t'}$.

$$m_i^t = F_m(\mathbf{u}_i^t, \mathbf{x}),$$
$$\mathbf{u}_i^{t'} = F_u(m_i^t).$$
Part A: Input-to-Label Message Passing

The first step passes messages from input $\mathbf{x}$ to each label embedding $\mathbf{u}_i^t$ to produce the intermediate embedding state $\mathbf{u}_i^{t'}$.

\[
\mathbf{m}_i^t = \Sigma_{j \in \mathbf{x}} F_m(\mathbf{u}_i^t, \mathbf{x}_i),
\]

\[
\mathbf{u}_i^{t'} = F_u(\mathbf{m}_i^t).
\]
Part B: Label-to-Label Message Passing

The second step passes messages between label embeddings to update their states conditioned on $\mathbf{x}$ to produce updated embedding $u_i^{t+1}$:

$$m_i^{t'} = \sum_{j \in \mathcal{N}(i)} F_m(u_i^{t'}, u_j^{t'})$$

$$u_i^{t+1} = F_u(m_i^{t'})$$
Readout Function

- After $T$ updates to the label embeddings, the last module predicts the probabilities of each label being positive $\{\hat{y}_1, \hat{y}_2, ..., \hat{y}_L\}$
- Readout function $R$ projects each of the $L$ label embeddings $\{u_{1:L}^T\}$ into a scalar value

$$\hat{y}_i = R(u_i^T; W^o) = \text{sigmoid}(W^o u_i^T)$$
Label Message Passing (LaMP) Networks

for T steps

Input-to-Label MP

Label-to-Label MP

$x$
Loss Function

Binary cross entropy, however since LaMP iteratively updates the label state for $T$ steps, we can impose a loss at each step $t$ by using the readout function to obtain intermediate predictions $\hat{y}_i^t = R(u_i^t; W^o)$

$$
Loss(y, \hat{y}^t) = \frac{1}{T} \sum_{t=0}^T \frac{1}{L} \sum_{i=1}^L - (y_i \log(\hat{y}_i^t) + (1 - y_i) \log(1 - \hat{y}_i^t))
$$
Label Graph Structure
Prior Label Graph Structure

- For datasets with known label structure (hierarchy, protein interaction), use known
- For all other datasets, we place an edge on the adjacency matrix for all labels that **co-occur in any sample** of the training set
Learned Label Graph Structure Using Attention

- Prior graphs may not be the best assumption since they are not conditioned on inputs
- We propose using attention to learn the graph end-to-end while training the classifier
Attention-Based MPNNs

\[ m_i^t = \sum_{j \in \mathcal{N}(i)} F_{m\alpha}(v_i^t, v_j^t), \]

\[ F_{m\alpha}(v_i^t, v_j^t) = \alpha_{ij}^t \mathbf{W}^\nu v_j^t \]
Attention-Based MPNNs

Attention coefficient $e_{ij}^t$ for pair of nodes $(v_i^t, v_j^t)$ is computed using some attention function producing a scalar value representing the relationship between nodes:

$$e_{ij}^t = (W_q v_i^t)^\top (W_u v_j^t)$$

Attention weight $\alpha_{ij}^t$ produced by normalizing over neighboring nodes

$$\alpha_{ij}^t = \text{softmax}_j(e_{ij}^t) = \frac{\exp(e_{ij}^t)}{\sum_{k \in \mathcal{N}(i)} \exp(e_{ik}^t)}.$$
Attention-Based MPNNs
Experiments and Results
Datasets

- We validate our model on 8 real world MLC datasets
- These datasets vary in the number of samples, number of labels, input type (sequential, tabular, graph, vector), and output type (unknown label graph, prior label graph)
LaMP Variations

LaMP$_{el}$ uses an edgeless label graph and messages are not passed between labels, assuming no label dependencies

LaMP$_{pr}$ uses a prior label graph where each label is able to attend to only other labels from the known label graph

LaMP$_{fc}$ uses a fully connected label graph where each label is able to attend to all other labels
## Example-Based F1 Scores (ebF1)

<table>
<thead>
<tr>
<th></th>
<th>Reuters</th>
<th>Bibtex</th>
<th>Bookmarks</th>
<th>Delicious</th>
<th>RCV1</th>
<th>TFBS</th>
<th>NUSWIDE</th>
<th>SIDER</th>
</tr>
</thead>
<tbody>
<tr>
<td>FastXML[2]</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.841</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Madjarov[3]</td>
<td>-</td>
<td>0.434</td>
<td>0.257</td>
<td>0.343</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SPEN[4]</td>
<td>-</td>
<td>0.422</td>
<td>0.344</td>
<td>0.375</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RNN Seq2Seq[5]</td>
<td>0.894</td>
<td>0.393</td>
<td>0.362</td>
<td>0.320</td>
<td><strong>0.890</strong></td>
<td>0.249</td>
<td>0.329</td>
<td>0.356</td>
</tr>
<tr>
<td>MLP</td>
<td>0.854</td>
<td>0.363</td>
<td>0.368</td>
<td>0.371</td>
<td>0.865</td>
<td>0.167</td>
<td>0.371</td>
<td><strong>0.766</strong></td>
</tr>
<tr>
<td>LaMP&lt;sub&gt;el&lt;/sub&gt;</td>
<td>0.883</td>
<td>0.435</td>
<td>0.375</td>
<td>0.369</td>
<td>0.887</td>
<td>0.310</td>
<td>0.376</td>
<td><strong>0.766</strong></td>
</tr>
<tr>
<td>LaMP&lt;sub&gt;pr&lt;/sub&gt;</td>
<td>0.902</td>
<td><strong>0.447</strong></td>
<td>0.386</td>
<td><strong>0.372</strong></td>
<td>0.887</td>
<td>0.321</td>
<td>0.372</td>
<td><strong>0.766</strong></td>
</tr>
<tr>
<td>LaMP&lt;sub&gt;fc&lt;/sub&gt;</td>
<td><strong>0.906</strong></td>
<td>0.445</td>
<td><strong>0.389</strong></td>
<td><strong>0.372</strong></td>
<td>0.889</td>
<td><strong>0.321</strong></td>
<td>0.376</td>
<td>0.764</td>
</tr>
</tbody>
</table>
## Speed

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reuters</td>
<td>0.788 (1.5x)</td>
<td>0.116 (2.1x)</td>
</tr>
<tr>
<td>Bibtex</td>
<td>0.376 (2.1x)</td>
<td>0.080 (2.1x)</td>
</tr>
<tr>
<td>Delicious</td>
<td>3.172 (1.1x)</td>
<td>0.473 (3.2x)</td>
</tr>
<tr>
<td>Bookmarks</td>
<td>9.664 (1.2x)</td>
<td>1.849 (1.3x)</td>
</tr>
<tr>
<td>RCV1</td>
<td>98.346 (1.2x)</td>
<td>1.003 (1.7x)</td>
</tr>
<tr>
<td>TFBS</td>
<td>187.14 (2.5x)</td>
<td>13.04 (4.2x)</td>
</tr>
<tr>
<td>NUS-WIDE</td>
<td>3.201 (1.2x)</td>
<td>0.921 (8.0x)</td>
</tr>
<tr>
<td>SIDER</td>
<td>0.027 (2.5x)</td>
<td>0.003 (21x)</td>
</tr>
</tbody>
</table>

Inside parenthesis shows the **speedup time** over LSTM probabilistic chain classifier.
Visualization: Intermediate Predictions
Visualization: Label to Label Attention
Conclusion

- We propose Label Message Passing (LaMP) Networks for multi-label classification
- LaMP networks model label interactions by placing labels as nodes on a graph
- LaMP networks are as accurate, faster, and more interpretable than the previous state-of-the-art MLC classifiers, as demonstrated on 8 datasets
Thank You

code available at: github.com/QData/LaMP

Arshdeep Sekhon

Dr. Yanjun Qi