Going Deep: A study in migrating existing analytics to use deep learning

Ryan Roser
@ryanroser
Director, Data Science & Text Analytics
Thomson Reuters Labs, San Francisco

OSCON 2018
An existing analytics product
Deep learning

Source: https://imgs.xkcd.com/comics/machine_learning.png
Our task is to transform a clock tower into a pile of linear algebra

Bag-of-words credit model

Deep Learning credit model
The StarMine Text Mining Credit Risk model launches in 2012

The model analyzes text for global public companies to estimate their forward 12-month default probability.
There is important foreshadowing hidden in the text.

### Comparison of Agency Rating and TR Model Components for Chesapeake Corp

<table>
<thead>
<tr>
<th>Agency Rating</th>
<th>TR Text Model Equivalent Rating</th>
<th>Stock Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td></td>
<td>$0–$24</td>
</tr>
<tr>
<td>AA</td>
<td></td>
<td>$3–$21</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>$6–$18</td>
</tr>
<tr>
<td>BBB</td>
<td></td>
<td>$9–$15</td>
</tr>
<tr>
<td>BB</td>
<td></td>
<td>$12–$12</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>$15–$15</td>
</tr>
<tr>
<td>CCC</td>
<td></td>
<td>$18–$15</td>
</tr>
<tr>
<td>CC</td>
<td></td>
<td>$21–$15</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>$24–$15</td>
</tr>
</tbody>
</table>

#### Transcript

**11/30/2007 – Transcript**

CEO: There is a reasonable possibility that we will not be able to comply with these covenants...

**08/31/2008 – 10-Q**

Based on current projections we are likely to not be in compliance with the financial covenants under the Credit required…. These matters raise substantial doubt about our ability to continue as a going concern.
The model has been a steady performer since its release
We have not updated the algorithm since it's initial release in 2012

TMCR outperforms the Altman Z-score, an industry standard benchmark

We see consistent performance on live data
Why migrate the model?

- Address six years of customer feedback
- Incorporate additional content & languages
- Forecast events on multiple horizons
- Improve overall performance
Artificial Intelligence is transforming every sector of the economy. How did we get here?

What is our approach?

Data
• Utilize domain knowledge from the prior model for features and enhancements
• Fit the deep learning model to the prior model’s default probability forecasts

Architecture
• The deep network design is informed by the structure of the prior model
Improve the training data using insights from prior work

Trim noisy data

- Exclude machine-generated news articles and broker research reports
- Exclude disclosure section of all broker research reports

Create features

- Apply age-weighting to documents → Include document date as a feature
- Apply text volume weighting → Include count of trailing docs as a feature
- Presence of certain sections in 8-K filings (current material event filings) are red flags
Train Deep Credit against the TMCR model predictions
Can deep learning generalize the existing model & outperform it?
Original Text Mining Credit Risk model

Documents (news, transcripts, research, filings)

“good morning… potentially violating the covenants…”

Stem words. Remove high frequency words. Chop into unique words and phrases

Bag-of-words

Potential, violat, the, covenant, potential violat, …

FULL Numerical representation of documents

<table>
<thead>
<tr>
<th></th>
<th>covenant</th>
<th>good morning</th>
<th>apple</th>
</tr>
</thead>
<tbody>
<tr>
<td>doc1</td>
<td>0</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>doc2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>doc3</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
</tbody>
</table>

Select most valuable words and phrases

Score docs

Aggregate doc scores

Learning Algorithm

Date Company Probability of Default

date1 A 0.0011
date1 B 0.0132
date1 C 0.0578
…

Make predictions

Document Scores

Date Company Probability of Default

date1 A 0.0011
date1 B 0.0132
date1 C 0.0578
…

Learn Algorithm

Select most valuable words and phrases

Score docs

Aggregate doc scores

Learning Algorithm

Date Company Probability of Default

date1 A 0.0011
date1 B 0.0132
date1 C 0.0578
…

Make predictions

Document Scores

Date Company Probability of Default

date1 A 0.0011
date1 B 0.0132
date1 C 0.0578
…

Learning Algorithm

Select most valuable words and phrases

Score docs

Aggregate doc scores

Learning Algorithm

Date Company Probability of Default

date1 A 0.0011
date1 B 0.0132
date1 C 0.0578
…

Make predictions

Document Scores

Date Company Probability of Default

date1 A 0.0011
date1 B 0.0132
date1 C 0.0578
…

Learning Algorithm

Select most valuable words and phrases

Score docs

Aggregate doc scores

Learning Algorithm

Date Company Probability of Default

date1 A 0.0011
date1 B 0.0132
date1 C 0.0578
…

Make predictions

Document Scores
**Data Engineering**

Documents (news, transcripts, research, filings)

“good morning… potentially violating the covenants…”

Stem words. Remove high frequency words. Chop into unique words and phrases

**FULL Numerical representation of documents**

<table>
<thead>
<tr>
<th></th>
<th>covenant</th>
<th>good morning</th>
<th>apple</th>
</tr>
</thead>
<tbody>
<tr>
<td>doc1</td>
<td>0</td>
<td>0</td>
<td>29</td>
</tr>
<tr>
<td>doc2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>doc3</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Select most valuable words and phrases

Score docs

Aggregate doc scores

<table>
<thead>
<tr>
<th>Date</th>
<th>Company</th>
<th>Probability of Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>date1</td>
<td>A</td>
<td>0.0011</td>
</tr>
<tr>
<td>date1</td>
<td>B</td>
<td>0.0132</td>
</tr>
<tr>
<td>date1</td>
<td>C</td>
<td>0.0578</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Make predictions

Learning Algorithm

Document Scores
We use a seq2seq architecture

- Sequence to Sequence models are often used for language translation e.g., they may translate a sequence of English words into a sequence of French words

Source: https://blog.keras.io/a-ten-minute-introduction-to-sequence-to-sequence-learning-in-keras.html
We use a seq2seq architecture

Instead of translating English to French, Deep Credit “translates” a sequence of documents into a sequence of probability of default (PD) forecasts.
Document Model

Document Encoder
INPUT: sequence of tokenized words from document

[The, company, has, submitted, a request, ... ]

- Embedding Layer
- CNN
- Document Encoding
- Dense Layer(s)
- Document Score

Encoder
INPUT: a sequence of tokenized document(s)

- Document Encoder
- GRU
- Internal Encoder State
- Document Encoding
Company Model

Decoder Input
Staring token or prior default probability

Encoder
INPUT: a sequence of tokenized document(s)

Dense Layer(s)

GRU

Encoder State

Document Model

Dense Layer(s)

Final output
Sequence of default probabilities for company
How well does it work?

Document encoder - CNN vs GRU

• CNN training time is faster (10x)
• Performance is similar (slightly lower for CNN)

Japanese text was supported by CNN document model

• Increase window size, increase document length
• AUC is comparable* to what we see for English: 0.79 AUC vs 0.86 for English

Company Model performance is nearly equivalent to TMCR
How do we interpret the model results?

• We’ve been thinking about interpretability from the outset

• The tests we do to confirm that the model works as intended are closely related to the types of visualization and outputs that users would like
Overall Score: 0.940

DEXTERA SURGICAL - FINAL SALE APPROVAL HEARING ANTICIPATED TO TAKE PLACE SHORTLY AFTER AUCTION OF CO’S ASSETS, ANTICIPATED CLOSING TO OCCUR BY EARLY 2018

Dec 12 (Reuters) - Dextera Surgical Inc DXTR.O:

DEXTERA SURGICAL FILES FOR CHAPTER 11 BANKRUPTCY AND SIGNS ASSET PURCHASE AGREEMENT WITH AESCULAP, INC.

DEXTERA SURGICAL INC - VOLUNTARY CHAPTER 11 PETITION WAS FILED IN UNITED STATES BANKRUPTCY COURT FOR DISTRICT OF DELAWARE

DEXTERA SURGICAL INC - ENTERED INTO AN ASSET PURCHASE AGREEMENT WITH AESCULAP INC, AN AFFILIATE OF B. BRAUN GROUP, FOR APPROXIMATELY $17.3 MILLION

DEXTERA SURGICAL - PROPOSED BIDDING PROCEDURES WOULD REQUIRE PARTIES TO SUBMIT COMPETITIVE BINDING OFFERS TO BUY CO’S ASSETS

DEXTERA SURGICAL INC - NEGOTIATED WITH AESCULAP FOR DEBTOR-IN-POSSESSION FINANCING

DEXTERA SURGICAL - FINAL SALE APPROVAL HEARING ANTICIPATED TO TAKE PLACE SHORTLY AFTER AUCTION OF CO’S ASSETS, ANTICIPATED CLOSING TO OCCUR BY EARLY 2018
Deep Credit forecasts a substantial rise in Parmalat’s default probability

UPDATE 1 - Parmalat lowers targets as economic outlook worsens

(Updates throughout with market reaction, background about previous targets)

MILAN, April 10 (Reuters) - Italian food group Parmalat <PRFI.MI> said on Thursday it was lowering forecasts for core earnings margins and expected its debt-equity ratio to be higher than expected amid a worsening economic climate. The company also said that it did not plan to issue bonds in the short term, two months after the company withdrew a planned bond offering amid concerns …
Deep Credit generalizes from TMCR and improves the document scores

<table>
<thead>
<tr>
<th>TMCR score</th>
<th>Deep Credit score</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>0.96</td>
</tr>
</tbody>
</table>

TMCR score: 0.95
Deep Credit Score: 0.97

BRIEF-$ORGANIZATION Unable To Return 1.5 Bln Yuan Due To Tight Liquidity, Large Debts
SAYS IT IS UNABLE TO RETURN A TOTAL OF 1.5 BILLION YUAN ($236.61 million) TO ACCOUNT FOR SHARE PRIVATE PLACEMENT DUE TO TIGHT LIQUIDITY, LARGE DEBTS

$ORGANIZATION SAYS IT IS UNABLE TO RETURN A TOTAL OF 1.5 BLN YUAN TO ACCOUNT FOR SHARE PRIVATE PLACEMENT DUE TO TIGHT LIQUIDITY, LARGE DEBTS
SAYS IT IS UNABLE TO RETURN A TOTAL OF 1.5 BILLION YUAN ($236.61 million) TO ACCOUNT FOR SHARE PRIVATE PLACEMENT DUE TO TIGHT LIQUIDITY, LARGE DEBTS
Wrapping up

Nine questions to consider before migrating to deep learning

1. What do you hope to gain by using deep learning?
2. Can deep learning achieve your objectives?
3. What insights and domain knowledge from the prior model can become features or improve training data?
4. How will you train the deep network?
5. Can your prior model be used to label training data?
6. Do you have the people you need?
7. How will you deploy to production?
8. What are the functional units of your prior model?
9. How can these function units be represented in a network?
Thank you!

For questions or more information:

Ryan Roser
@ryanroser
Director, Data Science & Text Analytics
Thomson Reuters Labs