BUILDING INTELLIGENT SOLUTIONS

Machine Learning & Natural Language Processing at Bloomberg

James Zhang, Ph.D.
Bloomberg Labs
First Bloomberg headline
NYT story
SEC announcement
Goal:

Data and Knowledge ➔
• Infer patterns
• Make predictions

Build/enhance INTELLIGENT solutions for our clients
Goal: Use Data and Knowledge about the real world in order to infer patterns and make predictions. Leverage this ability to build and enhance the tools our clients need.

Sentiment Analysis
Novelty Detection
Market Impact Analysis
Event Detection & Extraction

James Zhang, Ph.D.
Bloomberg Labs
Goal: Use Data and Knowledge about the real world in order to infer patterns and make predictions. Leverage this ability to build and enhance the tools our clients need.

Sentiment Analysis
Novelty Detection    Topic-based Clustering
Market Impact Analysis    Social Media Velocity
Event Detection & Extraction    Machine Translation
Social Graph Analysis
User Behaviour Analysis

James Zhang, Ph.D.
Bloomberg Labs
**Goal:** Use Data and Knowledge about the real world in order to infer patterns and make predictions. Leverage this ability to enhance the tools our clients need.

- Sentiment Analysis
- Novelty Detection
- Topic-based Clustering
- Market Impact Analysis
- Social Media Velocity
- Language Detection
- Event Detection & Extraction
- Machine Translation
- Social Graph Analysis
- Tokenization
- User Behaviour Analysis
- Parsing

James Zhang, Ph.D.
Bloomberg Labs
Tools:

- Machine Learning
- Natural Language Processing

Conclusion

Prediction

James Zhang, Ph.D.
Bloomberg Labs
Develop the fundamental tools that are required in order to represent and draw conclusions from our knowledge and data. Machine Learning and Natural Language Processing both play a part in this effort.

Tools:

- Supervised Learning
- Semi-supervised Learning
- Unsupervised Learning

James Zhang, Ph.D.
Bloomberg Labs
• **Features**
  - Discriminating characteristics
    - How ‘useful’ they are

• **Statistical models**
  - Features $\rightarrow$ decisions (similar to human-being)
  - **Classification**
    - Predict class assignment
  - **Regression**
    - Predict the value of an attribute.
The simplistic view from 10,000 feet:
The simplistic view from 10,000 feet:

How should we select features?

James Zhang, Ph.D.
Bloomberg Labs
One often used measure...

The reduction in uncertainty about one random variable given knowledge of another. **How dependent are two distributions on each other?** Especially useful as a scoring method during feature selection for Machine Learning.

Where $H(X)$ is the entropy of random variable $X$. 

James Zhang, Ph.D.
Bloomberg Labs
Depending upon the data, and the expert knowledge available, different modeling and inference techniques will be effective.

 Generative Models

 Naïve Bayes
 Hidden Markov Models
 Latent Dirichlet Allocation
 ...model the joint probability $P(x,y)$

 Discriminative Models

 Logistic Regression
 Conditional Random Fields
 Support Vector Machines
 ...model the conditional probability $P(y|x)$
Features
- Satellite image
- Oil storage tank levels

Model
- Linear regression

Value of interest
- Future oil price
- Classify documents
  - About cats / dogs
- Feature: ‘cat*’ / ‘dog*’
- Each point in the graph is a document
  - The position along an axis: number of mentions:

>cats are not dogs. This document is a bit biased towards cats. cats are great. cats have nine lives.

\[
\text{Cat}^* = 4, \text{Dog}^* = 1
\]
K(x_i, x_j) = \phi(x_i)\phi(x_j)

Margin = \frac{2}{\sqrt{w^Tw}}

\xi > 1

Support Vector

\xi < 1

Misclassified point

\xi = 0

w^T\phi(x) + b = -1

w^T\phi(x) + b = 0

w^T\phi(x) + b = +1

image courtesy of research.microsoft.com “The Standard SVM Formulation”

James Zhang, Ph.D.
Bloomberg Labs
• The Naïve Bayes classifier:
  • Predict an unseen instance with highest posterior probability
  • Assume each piece of evidence conditionally independent
  • Need sufficient training data

**Attribute**: Rain :: Lightning \(\rightarrow\) Thunder?

\[
P(\text{Lightning} & \text{Rain} | \text{Thunder}) = P(\text{Rain} | \text{Thunder}) \times P(\text{Lightning} | \text{Thunder})
\]

\[
C_{map} = \text{arg max}_{c \in C} \hat{P}(c | d) = \text{arg max}_{c \in C} \hat{P}(c) \prod_{1 \leq k \leq n_d} \hat{P}(d_k | c)
\]

\[
\sim \text{arg max}_{c \in C} [\log \hat{P}(c) + \sum_{1 \leq k \leq n_d} \log \hat{P}(d_k | c)]
\]

James Zhang, Ph.D.
Bloomberg Labs
Develop the fundamental resources that are required in order to represent and draw conclusions from our knowledge and data. Machine Learning, Statistical Measures, & all play a part in this effort.

**Tools:**

- Natural Language Processing
- Named-Entity Recognition
- Syntactic Parsing
- Named-Entity Resolution
- Tokenization
- Semantic Role Labeling
- Inference
- Stems
- Word Sense Disambiguation
- Language and Region Detection

Each one of us bought a cup of Java in Starbucks and we then started discussing the Java project from [CS430](#).
“James loves Machine Learning when the objective is well-specified.”
Mark the boundaries of tokens and sentences. Maximize the probability of each segment.

[“James loves Machine Learning when the objective is well-specified.”]

[“当目标明确时詹姆士喜欢机器学习领域.”]

[“jameslovesmachinelearningwhentheobjectiveiswellspecified.”]
Identify the grammatical categorization of each token:

[ James loves Machine Learning when the objective is well-specified . ]
Noun groupings and named-entity recognition provides identification of logically connected entities. Named-entity resolution ties those entities to an ontology (knowledge-base) to differentiate them:

[ James loves Machine Learning when the objective is well-specified. ]

The ontology may also be used to resolve definitions of ambiguous words:

Verb, Present Indicative: to hold dear: CHERISH

James Zhang, Ph.D.
Bloomberg Labs
Build the dependency relationship diagram between constituents of the sentence:

James loves Machine Learning when the objective is well-specified.
What are the roles of different constituents in the sentence?

[ James loves Machine Learning when the objective is well-specified. ]

James Zhang, Ph.D.
Bloomberg Labs
If we get this far, we can make our findings portable:

\[ \exists x \in X, \exists y \in Y : \text{hasObjective}(x,y) \land \text{wellSpecified}(y) \land \text{loves}(\text{James,y}) \]

[ James loves Machine Learning when the objective is well-specified. ]
Deliver: Apply the tools available in order to build higher-level solutions that leverage latent knowledge in the data, and expert intuition.

Sentiment Analysis
Novelty Detection  Topic-based Clustering
User Behaviour Analysis  Social Media Velocity  Language Detection
Market Impact Analysis  Machine Translation
Event Detection & Extraction
Social Graph Analysis

James Zhang, Ph.D.
Bloomberg Labs
### Sentiment Analysis

#### Motivation

James Zhang, Ph.D.

Bloomberg Labs

---

<table>
<thead>
<tr>
<th>Sentiment Daily Avg</th>
<th>SMAVG (5)</th>
<th>AAPL US Last Price</th>
<th>AAPL Rate of Change (5)</th>
<th>Correlation (E/G)</th>
<th>Signal</th>
<th>Good/Bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0012</td>
<td>0.0441</td>
<td>319.31</td>
<td>0.5542</td>
<td>0.421</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>0.0967</td>
<td>0.0512</td>
<td>316.00</td>
<td>2.871</td>
<td>-0.086</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.079</td>
<td>0.0701</td>
<td>324.34</td>
<td>3.1198</td>
<td>-0.595</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.0819</td>
<td>0.0714</td>
<td>321.55</td>
<td>3.1242</td>
<td>-0.072</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.0483</td>
<td>0.0558</td>
<td>322.40</td>
<td>3.9096</td>
<td>0.196</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.0353</td>
<td>0.0465</td>
<td>317.40</td>
<td>0.532</td>
<td>-0.457</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.0576</td>
<td>0.0605</td>
<td>311.09</td>
<td>-0.7053</td>
<td>0.277</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>0.0814</td>
<td>0.0655</td>
<td>316.78</td>
<td>3.3058</td>
<td>-0.431</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.0359</td>
<td>0.0543</td>
<td>315.72</td>
<td>2.3766</td>
<td>-0.436</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.0575</td>
<td>0.0486</td>
<td>314.73</td>
<td>-0.438</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0.0702</td>
<td>0.0375</td>
<td>314.72</td>
<td>4.7571</td>
<td>0.533</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.0828</td>
<td>0.0353</td>
<td>308.64</td>
<td>2.3675</td>
<td>0.478</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

---
Sentiment Analysis
Modeling Classes

James Zhang, Ph.D.
Bloomberg Labs
Market Impact Analysis
Predicting Events

James Zhang, Ph.D.
Bloomberg Labs
Machine Translation

James Zhang, Ph.D.
Bloomberg Labs
Bloomberg

Machine Translation

A mammoth task, showcasing a wide range of NLP and learning problems, applicable to most linguistic information extraction workflows.

Indexing, Classification, and Search

- Question Answering;
- Query routing and matching;
- Interlingua processing (Sentiment, Market Impact, Novelty, etc.);

End-user exposure

- On-demand news, communication, UI translation;
- Prevent cross-lingual information arbitrage;
- Translation as a service;

Architecture

- Massive parallelization and optimization;
- In-memory models can run into 10’s of GBytes;

James Zhang, Ph.D.
Bloomberg Labs
Frame translation as a conditional probability problem:

• \[ P(e|f) = \frac{P(f|e) \times P(e)}{P(f)} \]
  
  • \( P(e) \) is the prior probability of English;
  • \( P(f) \) the prior probability of foreign;
  • Find the English phrase with the maximum likelihood of being the intended phrase;

Priors are taken from monolingual Language Models and bi-lingual translation models;

Need large corpora of aligned words/phrases/sentences to approximate an optimal solution;
It’s difficult to reason directly from the $P(english|French)$ model:

- $P(f|e)$—bag of words approach?
  - Not a great model of the translation process:
    \[ P(\text{the, boy, runs}) = P(\text{runs, boy, the})! \]

- $P(e)$—high probability for grammatical sentences?
  - Difficult to achieve in practice;

- But, if we combine these two together:
  - $P(\text{the boy runs}) > P(\text{runs boy the})$
    \[ \rightarrow P(f|e) \times P(\text{the boy runs}) \text{ will be higher;} \]

- $P(e)$ worries about English word order so $P(f \mid e)$ doesn't have to;

- Similar models can be used to capture a rich set of cross-lingual relationships—with weighted combinations to optimize translation quality;
Centerra Gold (CG) today reported a net loss of $54.6 million, or $0.23 per share.
Finding the best translation in practice:

- Build an n-best list for each word/phrase;
- Perform a beam-search from an empty target sentence state, to a target sentence that minimises the cost of each transition, given the hypotheses generated;
- At each transition, prune the search space using cost until now, and future cost estimation;
Questions: