Predicting Employee Expertise for Talent Management in the Enterprise

Kush R. Varshney, Vijil Chenthamarakshana, Scott W. Fancher, Jun Wang, Dongping Fang, and Aleksandra Mojsilovic
Human Resource Analytics

The Waves of Business Analytics

- Finance & Logistics
  - Integrated ERP and Financial Analytics
- Customer & Marketing
  - Predictive Customer Behavior - CRM
- Talent & Leadership
  - Predictive Talent Models
  - HR Analytics
- Integrated Supply Chain
- Web Behavior Analytics
- Business-driven Talent analytics
- 1980s Financial and Budget Analytics
- Customer Segmentation Shopping Basket
- Integrated Talent Management Workforce Planning
- Customer Analytics – CRM (Data Warehouse)
- Recruiting, Learning, Performance Measurement
- Logistics and Supply Chain analytics

The Industrial Economy
- Steel, Oil, Railroads

The Financial Economy
- Conglomerates Financial Engineering

The Customer Economy and Web
- Customer Segmentation Personalized Products

The Talent Economy
- Globalization, Demographics Skills and Leadership Shortages

Early 1900s

1950s-60s

1970s-80s

Today

Image source: Bersin (2013)
Talent Analytics

• Largest worldwide employers today are knowledge-based enterprises

• Most important asset is human capital (Schultz, 1961)

• Knowledge workers are unique, each having individual skills and expertise

• Most basic of problems: *inventorying* employees according to expertise

image source: http://www.humancapitalstrategygrp.com/workforce-development
Why a Talent Inventory is Needed

• Quickening pace of technological innovation
  – New products, solutions, and acquisitions emerge each quarter

• Important for strategic and tactical business decision-making to be informed by complete, precise, accurate, and up-to-date information on the expertise of employees (Hu, Ray, and Singh, 2007)

• Tactical example: What team should serve a given client (in terms of composition of employee skills)

• Strategic example: Which emerging technology areas does the company have the talent to support
Problem Statement

• Develop predictive analytics based upon employees’ digital footprints to constantly update the current inventory of expertise across an organization in a way that commingles with existing business processes
Relationship to Prior Work

• LinkedIn and other similar skill recommendation systems have free-form skill description
  – Cannot integrate with existing ecosystems of processes and reporting tools built around expertise taxonomies

• Prior work on expertise prediction within an enterprise has been based only on internal social media data and has not integrated with business processes (Shami et al., 2009; Guy et al., 2013)
IBM Corporation

• Focus on a deployed system within the IBM Corporation

• Approximately 425,000 employees worldwide
  – Hardware, software, consulting services, research, sales, support, …

• Five-level expertise taxonomy
  – Sample values on next slide

• Employees assess themselves against the taxonomy
  – A significant fraction have incomplete, incorrect, or out-of-date assessments
## IBM Expertise Taxonomy

<table>
<thead>
<tr>
<th>Taxonomy Level</th>
<th>Sample 1</th>
<th>Sample 2</th>
<th>Sample 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Job Category</td>
<td>Sales</td>
<td>Human Resources</td>
<td>Research</td>
</tr>
<tr>
<td>Secondary Job Category</td>
<td>Industry Sales</td>
<td>Learning</td>
<td>Research Staff</td>
</tr>
<tr>
<td>Job Role</td>
<td>Brand Client Representative</td>
<td>Learning Consultant</td>
<td>Research Scientist</td>
</tr>
<tr>
<td>Job Role Specialty</td>
<td>Brand Client Representative: BAO-Advanced Analytics &amp; Optimization</td>
<td>Learning Consultant: Collaboration, Knowledge &amp; Communities</td>
<td>Research Scientist: Computational Biology</td>
</tr>
<tr>
<td>Skill</td>
<td>Sell ILOG Optimization</td>
<td>Analyze Performance Improvement Needs</td>
<td>Develop Algorithms for Biological Data Analysis</td>
</tr>
</tbody>
</table>
Machine Learning Formulation

• Treat job role or specialty prediction for an employee as a supervised classification problem
  – Very large number of classes

• Varied features derived from employees’ digital footprints within the company

• Present top $k$ predictions as output along with confidence value

• Misclassification error is in fact the most appropriate performance metric

• Experiment with one-versus-all $\ell_1$-regularized logistic regression, $\ell_2$-regularized logistic regression, SVM, naïve Bayes
Features

A. Employee-entered free text on their responsibilities

B. Basic HR information

C. Internal social media (tags, blogs, wikis, etc.)

D. Job-specific data sources like sales opportunities for salespeople, publications for researchers, etc.
Empirical Study

• Predicting job roles of salespeople

• 11 class problem (imbalanced)
  – Brand Client Representative (BCR), Client Representative (CR), Client Technical Architect (CTA), Client Technical Manager (CTM), Client Technical Specialist (CTS), Client Unit Executive (CUE), Industry Solution Representative (ISR), Mid-Market Client Representative (MCR), Solution Representative (SR), Solution Representative - Brand Specialist (SRB), and Solution Sales Manager (SSM)

• Approximately 37,000 employees in training set; 5,000 in test set
# Fivefold Cross-Validation Accuracy

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Job Title (A)</td>
<td>0.6746</td>
<td>0.6749</td>
<td>0.6695</td>
<td>0.6410</td>
</tr>
<tr>
<td>HR Info (B)</td>
<td>0.7661</td>
<td>0.7641</td>
<td>0.7604</td>
<td>0.6807</td>
</tr>
<tr>
<td>Social Tags (C)</td>
<td>0.2320</td>
<td>0.2396</td>
<td>0.2380</td>
<td>0.2573</td>
</tr>
<tr>
<td>Sales Opp (D)</td>
<td>0.3374</td>
<td>0.3404</td>
<td>0.3473</td>
<td>0.2306</td>
</tr>
<tr>
<td>(A) + (B)</td>
<td>0.8016</td>
<td>0.8031</td>
<td>0.7899</td>
<td>0.7330</td>
</tr>
<tr>
<td>(A) + (B) + (C)</td>
<td>0.7671</td>
<td>0.7703</td>
<td>0.7504</td>
<td>0.6118</td>
</tr>
<tr>
<td>(A) + (B) + (C) + (D)</td>
<td>0.7720</td>
<td>0.7733</td>
<td>0.7655</td>
<td>0.3952</td>
</tr>
</tbody>
</table>
Cross-Validation Accuracy ($\ell_2$-Regularized Logistic Regression)
Deployment
Impact

• Initial deployment (without interface) to obtain correct job roles for approximately 4,000 worldwide salespeople

• To get salespeople to correctly enter their expertise would have taken approximately 30 minutes per employee
  – Save 1 person-year of effort = $1M of revenue generated by salesperson

• Estimate 20 person-years of effort savings when deployed to entire company for annual assessments

• No need to limit to annual assessment
  – Just-in-time inventories, point-of-sale inventory updates, economic order quantities, and predictive inventory demand become possible for human resources and expertise management
Summary

• Talent and human capital is a knowledge-based company’s most valuable resource that must be harnessed properly using trusted expertise information

• Developed a classification methodology to predict the expertise of employees based on features derived from the digital footprints of employees
  – Label set from expertise taxonomy

• In the process of deploying the system for use by IBM Corporation
  – Should result in approximately twenty person-years of savings in annual updates of job roles and specialties
  – Impact is even greater than the savings in manual effort, because all business processes that depend on complete, accurate, and updated expertise data benefit from the predictions
  – Because of the steep reduction in effort, it will now be possible to update expertise assessments much more frequently than once a year, which is a transformation required to compete in today's dynamic business environment
Questions