Predicting and Recommending Skills in the Social Enterprise

Kush R. Varshney, Jun Wang, Aleksandra Mojsilović, Dongping Fang, and John H. Bauer
Business Analytics and Mathematical Sciences
IBM Thomas J. Watson Research Center
1101 Kitchawan Road
Yorktown Heights, New York 10598

Abstract

In this paper, we discuss the need for accurate skills assessments of employees in large, global, client-facing enterprises and shortcomings of existing systems for obtaining and managing expertise. We describe enterprise and social data that can be mined to improve skill assessment processes. We propose a matrix completion approach with side information for improved skill assessment prediction and recommendation, and discuss how outputs can improve existing business process use cases and illuminate new ones.

Introduction

Many large multinational corporations are transforming themselves into globally-integrated enterprises (Palmisano 2006). Simultaneously, they are requiring more of their employees to work directly with clients (Prahalad and Ramaswamy 2004). Moreover, the use of social networking and social media technologies inside the business is increasing (McAfee 2006). Due to these prevailing trends, there are many new management problems to tackle, many new data sources to utilize, and many new opportunities to transform how enterprises manage, engage, utilize, and plan for their workforce.

Knowledge workers are not interchangeable because they each have specialized expertise and skills; capturing that individual specialty is critically important for the successful operation of a large enterprise for many reasons. If skills of employees are properly catalogued, then the appropriate expert (from any business unit and region of the world) can be called upon to meet with a client, answer a question, and join a project staff. Skill information can be also used for management and planning on a more aggregate level.

Unfortunately, capturing skill information on employees is not easy. First of all, from a semantics and pragmatics point of view, it is difficult for two people (let alone tens of thousands of people within an enterprise) to agree on definitions and delineations of skills. Second, even if a common set of skills and definitions is decided upon, it is difficult to obtain accurate, validated assessments of all employees on all skills.

In this paper, we describe existing, so-called Enterprise 1.0, approaches to obtaining and managing skills information and discuss their shortcomings. We then suggest an improved approach that does not discard legacy systems, but rather enhances them through the use of structured and unstructured data made available by the new adoption of enterprise social media, as well as the use of structured and unstructured enterprise data that is often siloed away from expertise management systems. In particular, we propose the use of collaborative filtering/recommender system/matrix completion technology to abet the collection of accurate, validated employee skill assessments in various ways. We also discuss more advanced use cases that are enabled by such technology.

The work herein is similar to previous work that aims to responsibly utilize social data generated by employees to discern signals of expertise, but can be distinguished because our approach melds with legacy business processes and data sources (John and Seligmann 2006; Shami et al. 2009). We note that the machine learning problem we discuss here, skill prediction and recommendation, is different than social matching (Terveen and McDonald 2005; Ehrlich and Shami 2008). In social matching, the object is to organically recommend people to other people, where in this work we are attempting to predict the expertise of people, primarily to recommend unassessed skills. Our problem shares similarities with recommending products or media to consumers, but has differences due to the many supported use cases we describe later in the paper.

Characteristics of Enterprise Skill Taxonomies

Employees can be defined by organization charts and reporting chains, or they can be defined by the specific functions that they carry out. In the terminology of Ilgen and Hollenbeck (1991), these two views are *jobs* and *roles*, and both are useful for various business processes (Hu, Ray, and Singh 2007; Naveh et al. 2007). As such, it is important to capture these orthogonal structures when capturing skills information.

Skills taxonomies, in which the child elements are specific skills and competencies, i.e. roles, and parent elements are organized around jobs, are a common way to structure skill assessment. If an enterprise had skill levels validly assessed for all employees for all skills in their taxonomy, and the skills listed in the taxonomy were appropriate for the current functions of the enterprise, then a skills taxonomy would be extremely valuable. As conveyed by Christian Archambeau, Principal Director of Human Resources for the European Patent Office, “knowing what skills and capability you have in your workforce and employing a common skills taxonomy allows you to move people around and create flexibility and agility.”

Unfortunately, however, the status quo is that skills taxonomies are usually complex, inflexible behemoths that are cumbersome to expand with emerging new skills, and whose elements are difficult to populate and update for individual employees. Moreover, Ilgen and Hollenbeck (1991) warn that “the stable appearance of organizational structures often masks the arbitrary nature of the structures. Jobs and roles evolve over time as the result of the interaction of physical and social systems in the organization and often stabilize for very arbitrary reasons.”

More specifically, regarding obtaining skill level assessments for employees, a common technique is self-assessment. However, such an approach is fraught with poor understanding of cryptic skill descriptions by employees, differing interpretations of levels of expertise, non-compliance, and so on. On top of that, if a taxonomy contains a very large number of skills (as is common), it is difficult for a fully compliant, knowledgeable employee to even know which skills he or she should be assessing. An alternative approach—having managers assess their managees—has many of the same issues. These issues combined with the difficulty in keeping a skills taxonomy fresh and relevant, suggests that new approaches be considered.

Employee Data in the Enterprise

The enterprise maintains quite a bit of data about employees and skills that can be mined for skill prediction and recommendation. Some data elements are part of existing expertise management systems, and others exist in enterprise systems unrelated to expertise and need to be integrated. Additionally, businesses are now producing social data that has not existed previously, but is extremely valuable for understanding the expertise of employees.

Human resources (HR) data, such as an employee’s background, self-assessed skills and skill descriptions, has been a critical component of enterprise information warehouses. Besides these basic data elements, other more comprehensive data sources include annual performance evaluations, up-to-date curriculum vitae, labor claims, project reports, and so on. All these data elements can be integrated for use in automated skill assessment. In addition, nowadays, social technologies are being used inside corporations. For example, employees from different business units also belong to various online technical communities and frequently exchange ideas and thoughts through microblogs and wikis. Mining this dynamic social content can help identify employees’ expertise instantly. This is extremely important in high-technology industries since technical skills are changing and evolving, and skill assessment should be able to capture such changes.

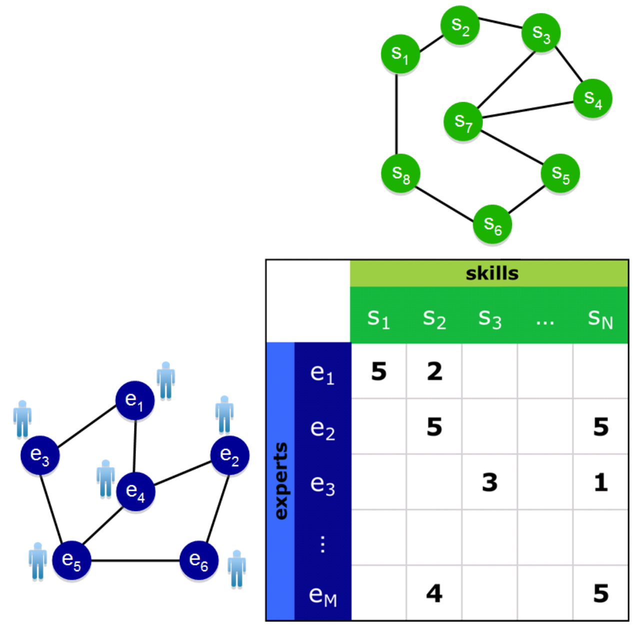

Figure 1: Illustration of matrix completion task for skill recommendation with skill-based and expert-based side information.

In particular, there are usually two types of information that can be extracted from social networking tools. The first type of data is the descriptive semantics, which is based on the online content that the employee read or posted. With natural language processing techniques, a set of keywords and a relevance weight vector can be associated with each employee. The second type of information from social technologies is relational data, including the co-membership of technical communities and co-authorship of wikis and project reports. Note that both the static HR data and the dynamic social media data can be treated as side information to help generate more accurate prediction of skills, as described below.

Skill Recommendation by Matrix Completion with Side Information

As illustrated in Figure 1, the expert-skill information can be treated as a matrix with the experts as rows and skills as columns. Given partially observed expert-skill data, the goal for skill recommendation is to predict the missing elements of the expert-skill matrix, namely *matrix completion*. In general, matrix completion is difficult and even theoretically impossible since missing entries could take any possible value. However, in many practical applications like the skill recommendation application, adding certain assumptions and additional side information can significantly help accomplish such a challenging prediction task.

Although the skill recommendation problem often involves tens of thousands of experts with thousands of possible skills, the number of hidden clusters of experts and skills could be very limited. In other words, experts and skills are often grouped in a small number of patterns. Mathematically, such a fact leads to a low-rank property of the expert-skill matrix.

Based on the low-rank assumption, many matrix factorization-based techniques are developed for various recommendation systems, ranging from Netflix's movie ranking to Amazon's product recommendation to Yahoo!'s online news suggestion (Koren, Bell, and Volinsky 2009). In addition, collaborative filtering is another category of closely related techniques for recommendation systems. It covers a wide range of techniques, including memory-based, model-based, and hybrid approaches. One of the clear advantages of collaborative filtering lies in its flexibility in integrating side information to improve the recommendation performance (Koren 2008; Chen et al. 2012).

In our skill recommendation problem, besides the partially observed expert-skill matrix, rich side information for either experts or skills can be accessed. For instance, through exploring HR data and the enterprise intranet, we can easily obtain the experts' information like organization information, community membership, microblog posts, etc. Similarly, the skill can be further assessed by using its description and definition. With such side information, the designed method not only improves the recommendation accuracy, but also can perform recommendations for new experts or skills. Such extremely difficult cases correspond to prediction of an entire missing row or entire missing column in the expert-skill matrix, which is unpredictable for conventional matrix completion techniques, as discussed by Wang, Varshney, and Mojsilović (2012).

There are a variety of evaluation methods and metrics available to quantify the performance of a matrix completion recommendation technique (Shani and Gunawardana 2011). Metrics include those based on ranking, root mean squared error, precision and recall, and normalized cumulative discounted gain. The process of testing can involve cross-validation and hold out sets, user simulations, and studies with real users. The evaluation choices depend on specific use cases for which the predictions are to be used. We discuss several use cases for the completed matrix and side information in the next section.

Use Cases

A reliable, automated skill representation and recommendation methodology, integrated with enterprise data, business processes, and collaborative/social technologies can be an essential element in how an enterprise manages its talent, plans its workforce, and enables its employees. Here we review several important use cases, ranging from collecting expertise information to finding the right experts.

Expertise Assessments

As already mentioned, manual skill assessments against large skill taxonomies suffer from the lack of compliance, lack of understanding and different interpretations, often yielding suboptimal and incomplete representation of expertise in the enterprise. A skill recommendation algorithm can be used to enhance and automate the expertise assessment process in multiple ways. First, rather than browsing the entire taxonomy, the recommender algorithm can pre-select the set of skills an employee is most likely to be expert on.

The algorithm can also remind the employees to re-evaluate their skills every time there is a significant discrepancy between the expertise assessment predictions and actual assessments. And most importantly, when integrated in enterprise collaboration tools or social platforms (imagine a pop-up box stating "Would you like to add Machine Learning to your profile?"), the recommender analytics make the skill assessment process seamless and transparent.

Expertise Endorsement

Expertise endorsement is another way to identify experts or quantify the level of expertise. The effectiveness of the endorsement scheme depends critically on how it selects experts to endorse, experts to conduct the endorsements, and the skills to be endorsed. The recommender algorithm can introduce a great deal of intelligence into this process, by identifying the experts who look similar in terms of both item features (similar skills) and user features (similar characteristics or close social proximity), thereby making sure that the endorsements are done by the users who know the expert and the expertise area well. One concern, especially if the endorser is revealed publicly, is that endorsement can transform into a communication medium or self-promotion medium rather than a means for improved skill assessment.

Skill Inventories

Understanding how many of critical skills there are in the organization, in a specific business unit, or a region is of tremendous importance for business planning ("we are launching a new product; do we have enough sellers with the required skill to generate planned revenue"), hiring ("how many sellers with skill X are we short of and need to hire"), and learning ("how many sellers have a related skill and can be quickly up-skilled into X"). The recommender algorithm can be a critical component towards fully automating and bringing new insight into strategic talent planning processes, as it provides the mechanism to automatically catalogue and count the relevant skills.

Taxonomy Management and Emerging Skills

In the manual setting, expanding the taxonomy with new skills and underlying assessments is cumbersome. As a result, it often takes several months or a year to conduct and collect the assessments from all employees. The recommender algorithm provides a much-needed automation and bootstrapping of the process; one can easily define a new skill via a collection of keywords (i.e. side information and item features), and then generate the skill assessment predictions based on item similarities and other information available in the system.

Finding Experts

Expertise location systems are quickly becoming an important business enablement tool. Sellers are looking for experts on a certain topic to answer clients question in real time.

Application developers are looking for colleagues who have worked in a certain area to ask a question on particular solutions, applications, or even a piece of code. New hires are looking for senior experts and technical leaders as mentors and role models. The ability to find just the right person, with the right expertise and characteristics will soon transform all people-centric processes in the organization, and take employee productivity, engagement and performance to the next level.

In its simplest form the recommender algorithm can be used as the expertise locator, to help employees find the top experts based on the specified skill ("Looking for an expert on healthcare analytics located in India to present to a client"). In a more advanced scenario, instead of a skill from the taxonomy, the seeker can enter a keyword or topic, and then narrow the search results based on region, brand, product, or job role. The seeker can even specify a query by providing a known expert as a template ("I am looking for an expert similar to John Smith III"); the algorithm can then return all the experts who are similar to John in both the user and item features.

Conclusion

Skill recommendation and prediction is an emerging problem of importance in large organizations that are increasingly globally-integrated, client-facing, and users of social technologies. As we have discussed, existing people-centric processes around skills and expertise, although having many merits, are not sufficient to support the future needs of a social workforce. Making use of the existing systems of expertise assessment and management, we propose an approach to enhance them for use in a social enterprise.

In particular, we propose enriching existing sparsely completed expertise assessment data with information captured about employees using enterprise social networking tools combined with HR and other management data, and with information captured about skills such as their descriptions and relationships to other skills. Then using factorization-based matrix completion approaches that incorporate side information about experts and skills, we propose providing predictions for all employees for all taxonomy skills.

Such predictions then enable many different use cases within the enterprise. Seamless and transparent completion of expertise assessments, endorsement of employees on skills by peers, skill inventorying for planning, taxonomy management under emerging specialties, and intelligent expertise location are some of the processes that can be fed by skill recommendation and prediction. We are currently in the process of developing, implementing, and deploying skill recommendation within IBM, a large enterprise with about one half million employees and twenty thousand skills in the existing taxonomy. Advanced technologies in the area of skills and expertise can truly facilitate the success of mobile, social, collaborative workforces.

Acknowledgments

The authors thank Donna Gresh, Nizar Lethif, and members of the Expertise Locator project team across IBM.

References

Chen, T.; Tang, L.; Liu, Q.; Yang, D.; Xie, S.; Cao, X.; Wu, C.; Yao, E.; Liu, Z.; Jiang, Z.; Chen, C.; Kong, W.; and Yu, Y. 2012. Combining factorization model and additive forest for collaborative followee recommendation. In *Proceedings of the KDD-Cup Workshop*.

Ehrlich, K., and Shami, N. S. 2008. Searching for expertise. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1093–1096.

Hu, J.; Ray, B. K.; and Singh, M. 2007. Statistical methods for automated generation of service engagement staffing plans. *IBM Journal of Research and Development* 51(3/4):281–293.

Ilgen, D. R., and Hollenbeck, J. R. 1991. The structure of work: Job design and roles. In Dunnette, M. D., and Hough, L. M., eds., *Handbook of Industrial and Organizational Psychology*. Palo Alto, CA: Consulting Psychologists Press. 165–207.

John, A., and Seligmann, D. 2006. Collaborative tagging and expertise in the enterprise. In *Proceedings of the WWW Collaborative Web Tagging Workshop*.

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *IEEE Computer* 42(8):30–37.

Koren, Y. 2008. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 426–434.

McAfee, A. P. 2006. Enterprise 2.0: The dawn of emergent collaboration. *MIT Sloan Management Review* 47(3):21–28.

Naveh, Y.; Richter, Y.; Altshuler, Y.; Gresh, D. L.; and Connors, D. P. 2007. Workforce optimization: Identification and assignment of professional workers using constraint programming. *IBM Journal of Research and Development* 51(3/4):263–279.

Palmisano, S. J. 2006. The globally integrated enterprise. *Foreign Affairs* 85(3):127–136.

Prahalad, C. K., and Ramaswamy, V. 2004. *The Future of Competition: Co-Creating Unique Value with Customers*. Boston, MA: Harvard Business School Press.

Shami, N. S.; Ehrlich, K.; Gay, G.; and Hancock, J. T. 2009. Making sense of strangers' expertise from signals in digital artifacts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 69–78.

Shani, G., and Gunawardana, A. 2011. Evaluating recommendation systems. In Ricci, F.; Rokach, L.; Shapira, B.; and Kantor, P. B., eds., *Recommender Systems Handbook*. New York, NY: Springer. 257–297.

Terveen, L., and McDonald, D. W. 2005. Social matching: A framework and research agenda. *ACM Transactions on Computer-Human Interaction* 12(3):401–434.

Wang, J.; Varshney, K. R.; and Mojsilović, A. 2012. Legislative prediction via random walks over a heterogeneous graph. In *Proceedings of the SIAM International Conference on Data Mining*, 1095–1106.