Optigrow: People Analytics for Job Transfers

Dennis Wei, Kush R. Varshney
Mathematical Sciences and Analytics
IBM Thomas J. Watson Research Center
Yorktown Heights, NY, USA
Email: {dwei,krvarshn}@us.ibm.com

Marcy Wagman
Workforce Management
IBM Global Technology Services
Poughkeepsie, NY, USA
Email: wagman@us.ibm.com

***Abstract*—The information technology (IT) services industry is undergoing a rapid change with the growth of market interest in cloud, analytics, mobile, social, and security technologies. For service providers to match this pace, they must rapidly transform their workforce in terms of job roles, and do so without incurring excessive cost while continuing to deliver core services. In this paper, we describe a big data approach to enable such a transformation through internal job transfers of suitable employees from legacy areas to growth areas. Toward this end, we use data on employee expertise to mathematically profile skill sets required for growth area jobs and develop a statistical scoring algorithm to prioritize internal candidates to be transferred to those growth area jobs. We describe how we have enacted this analytics procedure within the IT services division of the IBM Corporation and provide empirical results. We also discuss the lessons learned during the deployment, focusing mostly on organizational reasons preventing wide uptake.**

***Index Terms*—enterprise transformation; expertise analytics; human capital management; total variation distance; workforce analytics;**

I. Introduction

Many new technologies, including those supporting growth areas such as cloud computing, analytics, mobility, social computing, and security, require new skills that are constrained in the marketplace. Information technology (IT) service providers such as IBM are challenged with managing their workforce to meet an increasing demand for skills in new computing models and technologies, while keeping service rates competitive and managing a declining demand for skills in older technologies. Managing this transition well is a critical success factor for a services business [1].

One way to transform the workforce is by laying off people currently working on legacy technologies and hiring new people from external sources skilled in the growth area technologies. However such a choice incurs high costs: economic and human costs associated with layoffs, recruiting and onboarding costs for the new employees, and costs resulting from productivity losses as people adjust to a new employer. A better and more cost-effective way is to find and transfer employees currently working in legacy areas who already have the expertise to deliver services in a growth area or would be able to do so with a small amount of training [2].

The contribution of this paper is the development of a data analytics algorithm, built upon structured and unstructured data indicating employee expertise, that enables the internal transfer of people to growth areas in IT services companies with very large workforces. Further, we detail how we have implemented such an algorithm and approach for IBM's IT services division. Our experience with IBM has taught us several lessons; the main lesson is that internal transfer programs, when not associated with appropriate transformational support, can be unsuccessful despite the data and analytics performing well, due to organizational issues.

The transfer of people into growth areas relies on the existence of suitable people in the workforce. This condition is met within large enterprises such as IBM's IT services division because its employees possess human capital [3]. To be more specific, employees have accumulated skills, knowledge, competencies, and expertise that they might not be using in their current assigned job role. Additionally, people can gain skills and are quite motivated to do so, for both intrinsic and economic reasons. The general area of human resources (HR) and human capital management is currently experiencing a sea-change to being driven by data [4]. The internal job transfer solution we have proposed and deployed is another example of this transformation to workforce analytics.

Our proposed solution is different than existing prior work, such as [5], [6], because we are dealing with permanent transfers of employees to new business units, not temporary staffing assignments. Additionally, this solution is different as our ultimate aim is to identify not only individuals who may immediately have the correct skills, but also those who have prerequisite skills that will allow them to obtain the required new skills with a small amount of training. Moreover, as we discuss later, we do not require a requestor to describe the types of employees they are seeking since we have found that requestors can be unreliable in specifying what they seek. Our problem is also not a constrained optimization problem that aims to schedule employee teaming as in previous work because of the permanent nature of the transfers; our problem is one of statistical profiling and scoring.

The remainder of the paper is organized as follows. In Section II, we describe the structured and unstructured data sources that we use in developing our people analytics solution. In Section III, we describe the mathematical formulation for the proposed algorithm. In Section IV, we discuss empirical results, impact, and lessons learned. Section V summarizes the work and suggests directions for future research.

978-1-4673-7278-7/15 $31.00 © 2015 IEEE
DOI 10.1109/BigDataCongress.2015.84

II. Data Sources

The analytics approach for internal job transfers relies on one main data source: expertise assessments (EA), and three other data sources: curricula vitae, project tracking data, and basic HR information. We detail these four data sources in this section in preparation for describing the proposed analytics algorithm.

Skills taxonomies, in which the child elements are specific skills and competencies, and parent elements are organized around jobs, are a common way for organizations to structure skill assessment [7], [8]. In this vein, IBM maintains an elaborate taxonomy of job categories, job roles, and skills corporation-wide, including in the IT services division. Moreover, IBM maintains assessments of employees against the taxonomy elements on a four point scale. Historically, these assessments have been self-made by the employees, but self-assessment is fraught with several issues [8], [9]; therefore, a new data-driven approach has been developed and deployed for EA, resulting in reliable data [9].

More specifically, IBM has a taxonomy of employee expertise with the following five coarse-to-fine levels: primary job category, secondary job category, job role, job role specialty, and skill. The taxonomy is a directed acyclic graph with parent-child relationships between values at different levels. We provide three examples of paths through the taxonomy with the five different levels separated by the greater than symbol: Sales > Industry Sales > Brand Client Representative > Brand Client Representative: BAO-Advanced Analytics & Optimization > Sell ILOG Optimization; Human Resources > Learning > Learning Consultant > Learning Consultant: Collaboration, Knowledge & Communities > Analyze Performance Improvement Needs; Research > Research Staff > Research Scientist > Research Scientist: Computational Biology > Develop Algorithms for Biological Data Analysis.

An individual employee is unassessed or has a zero-valued assessment for most items in the taxonomy, but has non-zero assessments for a handful of job roles and job role specialties. One job role and one job role specialty is usually designated as the primary one for the employee. EA provides the key data points from which to estimate expertise profiles for jobs in growth areas and from which to score employees against those profiles.

The second and third data sources, curricula vitae and claims, also provide information on employee skills and expertise. Nearly all IBM IT services delivery employees have an unstructured curriculum vitae (CV) used both internally and externally. Since these CVs usually represent the skills and experience for which the employee is currently deployed, they can provide only supplemental expertise information rather than be the primary basis for a people analytics solution. As part of their regular job responsibilities, a certain subset of delivery personnel bill claims for their services. Specifically, the hours worked, the detailed service provided as encoded in a product and service taxonomy, and the client are recorded. The time reporting data very directly indicates what an employee is working on and thus has expertise to work on, but does not show what else the employee could do if given the opportunity.

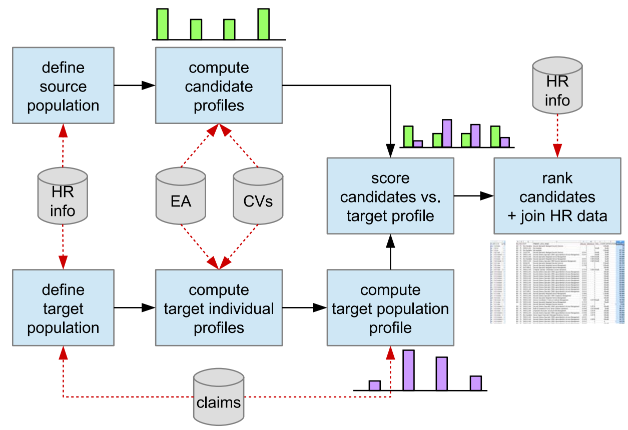

Fig. 1. Block diagram of proposed analytics algorithm. “EA” stands for IBM’s expertise assessment database or a similar source of expertise information.

Finally, basic HR information, such as the employee’s work location; the group, department, and business unit to which the employee belongs; and the employee’s pay grade is useful information around which to structure the proposed solution. The set of employees in a group or department is needed to delineate whom to base a job profile upon. Work location, business unit membership, and pay grade are all used to set constraints on the population of employees from which to draw candidates for internal transfers.

III. Analytics Algorithm

The proposed analytics algorithm evaluates candidates from a specified *source* population against an expertise profile constructed to be representative of a *target* population. Fig. 1 gives a pictorial overview of the algorithm. Individual steps are discussed in the following subsections: In Section III-A, details are given on the definitions of the source and target populations. Section III-B discusses the mathematical representation of the expertise of individuals based on the data sources outlined in Section II. From these individual representations, a profile for the target population is created as described in Section III-C. Section III-D discusses the scoring of candidate expertise profiles with respect to the target profile and the presentation of results.

A. Target and Source Populations

A distinguishing feature of our approach to job transfers is that the target for the transfer is defined by a population rather than a single job title or job role specialty. This target population can be designated in several ways. The most common way that we have used is to identify the hiring manager for the position together with his/her direct reports — in other words, the team that the successful candidate would join. Other teams that perform similar functions, possibly up to an entire department, can be included to increase the size and breadth of the target population. Alternatively, the target population may be defined by one or more service offerings and consist of all employees who have recently provided these services (for example within the last year).

The source population includes all employees that the company may wish to consider for a job transfer, and as such is potentially quite broad. To simplify some business procedures, we have usually restricted it to the same business unit to which the target population belongs. We have also excluded employees who can be identified as already working in a growth area so as not to detract from the overall workforce reallocation strategy (the challenge of excluding growth areas is discussed further in Section IV-D). In addition, the source population is often constrained by geography and pay grade. For jobs that require physical presence at a facility or co-location with team members, the source pool may be limited to the same city or state/province. Even if this is not the case, restrictions to the same country may be made for legal, language, and other logistical reasons. The pay grades of candidates are typically constrained to be the same or one lower than the pay grade associated with the open position. This reflects the fact that an employee is unlikely to accept a pay cut or to be prepared for a large increase in pay grade and responsibilities. Similarly, the few employees with senior managerial or executive job roles may be excluded.

B. Representation of Expertise

We consider multiple measures of expertise selected from the data sources in Section II. In what follows, we focus on the three lowest levels of the IBM expertise taxonomy: job role, job role specialty, and skill, as well as CVs. The two highest levels of the taxonomy, primary and secondary job category, are not included because they are too generic, as can be seen from the examples in Section II. On the other hand, professional or industry certifications, an additional data source that is sometimes available, can be handled similarly to skills.

An employee's expertise is characterized by a profile consisting of four non-negative vectors $\mathbf{x}^{\mathrm{JR}}$, $\mathbf{x}^{\mathrm{JRS}}$, $\mathbf{x}^{\mathrm{S}}$, $\mathbf{x}^{\mathrm{CV}}$, one for each of the expertise measures. For job roles and job role specialties, each entry x_j^{JR} or x_j^{JRS} represents the employee's capability in job role or specialty j. The total number of entries, i.e. the dimension of the vector, is determined by the number of job roles or specialties held by at least one individual in the target and source populations. Representation of skills and CVs is discussed later in this subsection.

We use a number of assumptions to assign values to the job role and job role specialty vectors $\mathbf{x}^{\mathrm{JR}}$ and $\mathbf{x}^{\mathrm{JRS}}$. First, since these two measures refer mostly to employee functions, for example selling a brand of analytics software or conducting research, we assume that 1) an employee's expertise in a job role or specialty is proportional to the time spent carrying out that function. By this logic, the sum of an employee's expertise in all job roles or specialties should be the same as for any other full-time employee working the same hours (we do not consider overtime for those employees to which it applies). Mathematically, this implies that the entries of $\mathbf{x}^{\mathrm{JR}}$ and $\mathbf{x}^{\mathrm{JRS}}$ sum to a constant, $\sum_j x_j^{\mathrm{JR}} = \sum_j x_j^{\mathrm{JRS}} = C$, which we take to be 1 without loss of generality. Hence x_j^{JR} can be interpreted as the fraction of time spent in job role j.

The remaining three assumptions are as follows: 2) For a given employee, all job roles and specialties designated as primary (of which there may be more than one) have equal weight, as do all non-primary (i.e. secondary) job roles and specialties. 3) Expertise in a primary job role or specialty is never less than that in a secondary one, i.e. $x_j \geq x_k$ for j primary and k secondary. 4) An employee spends at least 50% of her/his time on primary job roles and specialties. The last assumption is a general guideline in IBM's IT services division. In making Assumption 2, we ignore the employee self-assessments discussed in Section II, which may be unreliable. Alternatively, it would be straightforward to devise a weighting scheme that incorporates the self-assessments.

Assumptions 1–4 completely specify the values of $\mathbf{x}^{\mathrm{JR}}$ and $\mathbf{x}^{\mathrm{JRS}}$. For an employee with N_p primary and N_s secondary job roles or specialties,

$$x_j = \begin{cases} \max\left\{\frac{1}{2N_p}, \frac{1}{N_p+N_s}\right\}, & j \text{ primary}, \\ \min\left\{\frac{1}{2N_s}, \frac{1}{N_p+N_s}\right\}, & j \text{ secondary}. \end{cases}$$

We now turn to the representation of skills and CVs. This requires a different approach, most clearly for CVs since the data consists of unstructured text. For skills, even though they are a structured data type within the IBM taxonomy, the total number of skills is much larger than the numbers of job roles or specialties. Furthermore, many skills do not help in distinguishing qualified candidates and thus their inclusion would needlessly increase computational requirements.

We employ two measures to capture the information in skills and CVs. First, we define a small number of keywords that are relevant to the position and restrict attention to skills and CVs containing at least one of the keywords in their text. As an example, for positions in the growth area of security, keywords might include more specific sub-areas such as "managed security services", "network security", "crisis management", or product/brand names such as "QRadar", "Cisco", "Checkpoint". These keywords can be supplied by the hiring manager or knowledgeable HR professional, or manually extracted from a text description of the position. (Automatic keyword extraction using natural language processing techniques would be an interesting subject for future study.) We then associate each keyword with a component j of the skill and CV vectors $\mathbf{x}^{\mathrm{S}}$ and $\mathbf{x}^{\mathrm{CV}}$. Second, we additionally consider those skills that, in the IBM taxonomy, are children of the job role specialties that are most prevalent in the target population. These job role specialties may be selected by thresholding the target population vector $\mathbf{y}^{\mathrm{JRS}}$ to be discussed in Section III-C, or by other criteria. Each of the job role specialties is also associated with a component j of the skill vector $\mathbf{x}^{\mathrm{S}}$. Thus the dimension of $\mathbf{x}^{\mathrm{S}}$ is equal to the number of chosen job role specialties plus the number of keywords, while the dimension of $\mathbf{x}^{\mathrm{CV}}$ is equal to the number of keywords alone.

Unlike with job roles and specialties, we do not assume that an employee's expertise pertaining to a skill or CV keyword/specialty is proportional to working time; in other words, we do not make Assumption 1. Instead we take the viewpoint that skills and CVs tend to represent domain knowledge that an employee has accumulated, rather than functions that are regularly performed. (This boundary is subjective however and other interpretations are reasonable.) Consequently, the vectors $\mathbf{x}^{\mathrm{S}}$ and $\mathbf{x}^{\mathrm{CV}}$ are not normalized to sum to a constant. Furthermore, since there is no primary/secondary distinction for skills and CVs, the previous Assumptions 3 and 4 do not apply. We do however continue with an analogue of Assumption 2, namely that all skills containing a given keyword or associated with a job role specialty are equivalent. In doing so we again ignore employee self-assessments of their skill levels.

Given the above assumptions, for each employee we define skill component x_j^{S} to be the number of skills in the IBM taxonomy that contain keyword j or are children of job role specialty j. We let $x_j^{\mathrm{CV}} = 1$ if the employee's CV contains any occurrence of keyword j (regardless of the number of occurrences) and $x_j^{\mathrm{CV}} = 0$ otherwise. Thus $\mathbf{x}^{\mathrm{S}}$ is a vector of non-negative integers while $\mathbf{x}^{\mathrm{CV}}$ is a binary vector.

C. Target Population Profile

The set of individuals in the target population is further summarized by a single expertise profile as indicated in Fig. 1. Mathematically, the problem can be seen as determining a representative for a set of objects, in this case the individual expertise profiles defined in Section III-B, to optimize an error criterion. For computational simplicity, in this work we use the arithmetic mean, applied separately to the four expertise vectors, and possibly including weights as discussed in the next paragraph. The resulting mean vectors are denoted as $\mathbf{y}^{\mathrm{JR}}, \mathbf{y}^{\mathrm{JRS}}, \mathbf{y}^{\mathrm{S}}, \mathbf{y}^{\mathrm{CV}}$. For job role and job role specialty vectors $\mathbf{x}^{\mathrm{JR}}$, $\mathbf{x}^{\mathrm{JRS}}$, the mean also has the advantage of being similarly normalized to sum to 1. In terms of error criterion, the mean minimizes the sum of the squared Euclidean distances between itself and the individual vectors. A more self-consistent but computationally demanding alternative would be to minimize the distance metric (2) to be introduced in Section III-D instead of Euclidean distance.

When the target population corresponds to a team or multiple teams, we use unweighted averages of expertise vectors over the team members to form the target profile. However, when the target population is defined by service offerings, we use a weighted average in which employees who have delivered the services in the past are weighted by the number of hours they have spent in doing so. These hours are obtained from the time reporting data discussed in Section II by filtering based on the target services, summing the hours claimed by each employee, and then linking these to the employee expertise data.

D. Candidate Scoring and Ranking

The steps described in the previous subsections produce an expertise profile $\mathbf{x}^{\mathrm{JR}}$, $\mathbf{x}^{\mathrm{JRS}}$, $\mathbf{x}^{\mathrm{S}}$, $\mathbf{x}^{\mathrm{CV}}$ for each individual in the source and target populations as well as an average profile $\mathbf{y}^{\mathrm{JR}}$, $\mathbf{y}^{\mathrm{JRS}}$, $\mathbf{y}^{\mathrm{S}}$, $\mathbf{y}^{\mathrm{CV}}$ for the target population. Candidates from the source population (represented by vectors $\mathbf{x}$) are then scored on how closely they match the average target profile ($\mathbf{y}$) according to the following weighted formula:

$$\begin{aligned} s = \alpha^{\mathrm{JR}} S\big(\mathbf{x}^{\mathrm{JR}}, \mathbf{y}^{\mathrm{JR}}\big) + \alpha^{\mathrm{JRS}} S\big(\mathbf{x}^{\mathrm{JRS}}, \mathbf{y}^{\mathrm{JRS}}\big) \\ + \alpha^{\mathrm{S}} S\big(\mathbf{x}^{\mathrm{S}}, \mathbf{y}^{\mathrm{S}}\big) + \alpha^{\mathrm{CV}} S\big(\mathbf{x}^{\mathrm{CV}}, \mathbf{y}^{\mathrm{CV}}\big). \end{aligned} \quad (1)$$

The scoring function $S(\mathbf{x}, \mathbf{y}) \in [0, 1]$ is given by (4) and is discussed further below. A typical choice of weights might be $\alpha^{\mathrm{JR}} = 12.5$, $\alpha^{\mathrm{JRS}} = 25$, $\alpha^{\mathrm{S}} = 37.5$, $\alpha^{\mathrm{CV}} = 25$, which sum to 100 for ease of interpretation. Job roles receive less weight than job role specialties because they are less specific. However, since job role specialty data is more likely to be missing or inaccurate, the weight for job roles should not be too small either. Along similar lines, while CVs are a valuable source of information, they are less frequently available compared to skills and hence are given lower weight.

The scoring function $S(\mathbf{x}, \mathbf{y})$ in (1) is in turn defined by a function $D(\mathbf{x}, \mathbf{y})$ that measures the distance between expertise vectors $\mathbf{x}$ and $\mathbf{y}$, as given below:

$$D(\mathbf{x}, \mathbf{y}) = \sum_j w_j \frac{(y_j - x_j)_+}{y_j}, \quad (2)$$

where $\{w_j\}$ is a set of weights that sum to 1 and $(z)_+ = \max\{z, 0\}$ denotes the positive part of z. The motivation behind the form of $D(\mathbf{x}, \mathbf{y})$ is to provide a qualitative indication of the cost required for a candidate to acquire additional necessary expertise. More specifically, $(y_j - x_j)_+$ represents the amount of expertise of type j that the candidate must gain with respect to the target. Acquiring this expertise generally requires resources, whether spent on training and courses or time on-the-job ramping up productivity. The weights w_j represent the relative importance and acquisition cost of each type of expertise. The cost is zero if the candidate already has the desired expertise, i.e. if $x_j \geq y_j$.

For job roles and job role specialties, the weights w_j in (2) are set equal to y_j, i.e., the importance of a job role or specialty is proportional to the fraction of time spent on it by an idealized target employee. Furthermore, since $\mathbf{x}$ and $\mathbf{y}$ are both normalized so that they have the same sum, $\sum_j x_j = \sum_j y_j = 1$, (2) becomes

$$\begin{aligned} D(\mathbf{x}, \mathbf{y}) &= \sum_j (y_j - x_j)_+ = \sum_j (x_j - y_j)_+ \\ &= \frac{1}{2} \sum_j |y_j - x_j| \, . \end{aligned}$$

The last quantity is the total variation distance [10] between $\mathbf{x}$ and $\mathbf{y}$ when regarded as probability mass functions.

For skills, y_j represents the average number of skills over the target population associated with keyword or job role specialty j, while for CVs, y_j is the average incidence of keyword j. These may be less indicative of importance than fractions of working time. For example, the IBM taxonomy may include many more skills containing one keyword or under one job role specialty than for another. For this reason we usually do not set $w_j = y_j$ for skills and CVs, instead using uniform weights or weights chosen by the hiring manager or HR.

The functional form in (2) implies that $D(\mathbf{x}, \mathbf{y}) \in [0, 1]$ regardless of input vectors $\mathbf{x}$ and $\mathbf{y}$ provided that $\sum_j w_j = 1$. Thus it would be reasonable to forgo further normalization and directly apply a weighted average as in (1) to the job role, job role specialty, skill, and CV distances. However, we have found in practice that some of the expertise measures may not utilize the full unit interval. This happens because the target expertise vector $\mathbf{y}$ is an average over many individuals and no single employee may come close to matching the breadth in $\mathbf{y}$, i.e., the smallest observed value of $D(\mathbf{x}, \mathbf{y})$ may be significantly greater than zero. To correct for this bias, we compute distances $D(\mathbf{x}, \mathbf{y})$ not only for all candidates in the source population, but also for all members of the target population to provide a comparison, in effect treating target employees as examples of well-qualified candidates. In computing $D(\mathbf{x}, \mathbf{y})$ for a target employee, the mean $\mathbf{y}$ may be modified to leave out the employee being evaluated. Given distance values for all target employees, we first take the complement to yield a score $1 - D(\mathbf{x}, \mathbf{y})$ that is higher for better matches as opposed to being lower. We then compute the p-quantile to obtain a single benchmark,

$$S_p(\mathbf{y}) = Q\left(\{1 - D(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in \text{target}\}; p\right). \qquad (3)$$

Typically we choose the 95th percentile, i.e. $p = 0.95$, which corresponds to the best-scoring target employees while being more robust to outliers than the maximum ($p = 1$). The target population quantile in (3) is used to further normalize candidate scores as follows:

$$S(\mathbf{x}, \mathbf{y}) = \frac{\min\{1 - D(\mathbf{x}, \mathbf{y}), S_p(\mathbf{y})\}}{S_p(\mathbf{y})}. \qquad (4)$$

The resulting scores cover more of the unit interval than the raw scores $1 - D(\mathbf{x}, \mathbf{y})$. This completes the definition of the scoring function $S(\mathbf{x}, \mathbf{y})$ in (1).

The final step is to produce a list of candidates ranked in decreasing order of their overall scores (1). The component scores for different expertise measures in (1) may also be shown for better understanding of a candidate's overall score. As an additional reference point, a statistic of the overall scores for target individuals (again treated as if they were candidates) may be computed, for example the median. In addition to numerical scores and employee identifiers, the list may include helpful HR information such as manager, department, and other organizational details, current pay grade, and geographic location. All of this information is presented to human resources staff, workforce and resource management teams, and the hiring managers for review.

IV. Results and Impact

A. Empirical Results

In this section, we summarize the results that have been obtained to date within IBM using the proposed analytics approach to internal job transfers.

First we discuss the case of a large team in a European country delivering services in one subfield of one of the growth areas mentioned in Section I. This team was one of the first to which the algorithm was applied. The corresponding source population consisted of all employees in that country working in the IT services division but not in the growth area. Four expertise measures were evaluated: job roles, job role specialties, skills, and professional certifications, the last treated similarly to skills and taking the place of CVs above (CV summaries only became available to us in later campaigns). Scores were computed both for candidates as well as for existing members of the team, the latter scores helping to normalize the former as described in Section III-D.

The calibration provided by the target population is illustrated in Fig. 2, which shows a histogram of the overall scores (1) for the team, obtained as an average of the scores for the four expertise measures with weights $\alpha^{\text{JR}} = 100/7$, $\alpha^{\text{JRS}} = 200/7$, $\alpha^{\text{S}} = 200/7$, and $\alpha^{\text{cert}} = 200/7$. Most team members score highly, with two-thirds above a score of 80 and almost one-third above 95. This concentration of scores increases confidence in the ability of the scoring method to identify similarly-skilled candidates. On the other hand, the histogram also shows two outlier individuals with low scores, a result of having job roles, specialties, and skills very unlike the majority of the team. As most teams are not completely homogeneous, the presence of such outliers is perhaps unavoidable. Furthermore, if we view the job transfer problem as one of classification, then the results in Fig. 2 can be seen as a form of cross-validation, but only for the positive class of qualified employees. Similar validation for the negative class is made difficult by the lack of a pure sample — if the larger IT services population contains some qualified candidates as hoped, it is by definition impure. The problem of classification with impure samples or "label noise" is discussed extensively in [11], [12].

From the scores computed for the source population, a list of the top 125 candidates was compiled and reviewed by the hiring managers and human resources in the country. The conclusions of the review are presented in Fig. 3. Perhaps the strongest validation of our algorithm is represented by the 10 candidates who, unknown to us at the time, had either been previously approached about joining the team but declined, were recent hires to the team, or were former team members. (The inclusion of recent hires can be explained by timing issues from the large number of ongoing transitions in a large population as reflected in IBM's HR and expertise databases.) 34 of the candidates were deemed promising enough to have their CVs retrieved manually (complete, formatted versions, not the unstructured summaries that later became accessible to us). Of those, 13 were determined to be suitable candidates

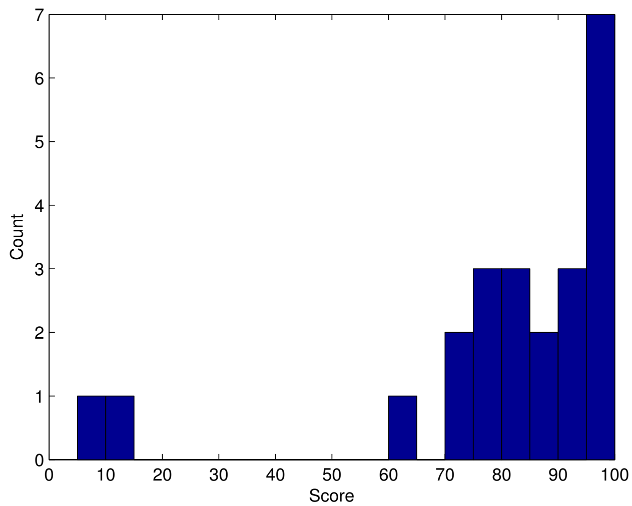

Fig. 2. Histogram of overall expertise scores for a team in Europe evaluated against a target profile representing itself. A large majority of scores are high as expected, thus validating the scoring method.

Fig. 3. Breakdowns of the internal candidates provided to the teams in (a) Europe and (b) Latin America.

for interviews. No further action was taken however as this evaluation was intended mainly as a test of the proposed algorithm. On the negative side, only 14 candidates were ruled out entirely as having unsuitable expertise, while the remainder were given neutral assessments, neither promising nor unsuitable.

Fig. 3 also shows a breakdown of candidates for a similar team in a Latin American country. In this case, the top 50 candidates were chosen from a source population consisting of all IT services employees in the country from outside the growth area. The same four expertise measures were considered. The results are broadly similar to those for the European country. Furthermore, the hiring manager in Latin America, motivated by an open position in his team, took the additional step of interviewing 7 of the candidates whose CVs were reviewed, and selected one to fill the opening. However, the manager at the time of the chosen candidate was reluctant to release the candidate because there were no policies or procedures in place to find a replacement. These organizational barriers are discussed further in Section IV-C.

Since the evaluations in the two countries, we have refined and applied the algorithm to about half a dozen batches of open positions, in total providing over 2300 ranked candidates for at least 90 positions and possibly many more (the exact number of positions was often not communicated to us). The feedback that we have received has generally been positive. However, to date we are not aware of actual transfers that have been completed as a result of our approach. The reasons for hesitation are discussed in Section IV-C. The algorithm has also provided valuable negative results: in one case, the apparent lack of suitable candidates gave human resources managers the confidence that proceeding with external hires was the right decision.

B. Projected Business Impact

As part of building the business case for this work, we estimated the monetary impact that the proposed algorithm could provide. Specifically, the benefit of an internal transfer results from the avoidance of costs associated with a layoff and the avoidance of costs associated with external hiring, namely recruiting and onboarding costs as well as initial productivity losses. Moreover, hiring replacement employees incurs a salary premium cost due to demand in the external hiring market [13].

To estimate these costs, from various departments within IBM we collected layoff cost data on a country-by-country basis as well as hiring cost and salary premium data at a finer level: country, job role, and pay grade. We then collected data on the distribution of employees in the IT services division by country, job role, and pay grade. Putting these pieces of data together through a weighted average, we obtained a savings estimate of over $50,000$ U.S. dollars per internal transfer not even including productivity losses. Most of this amount was due to layoff costs. The conclusion is that even a small number of enabled internal transfers would represent a large benefit for a company.

C. Organizational Challenges to Adoption

The identification of potential candidates through the Optigrow process was conceptually accepted and supported in the organization. The logic of the process and the business benefits were all well understood and supported by management. However, in the specific instance of identifying and moving a real person, the required collaboration and compromise by the hiring manager and current manager proved to be daunting and a significant inhibitor to achieving the expected business benefits.

While there are very strong incentives, primarily through utilization measurements, for IT services managers to 'lend' their team members to projects or other areas of the business, there is a very strong bias against moving individuals on a more permanent basis to a new job or reporting structure. The factors that contribute to this bias include the following:

1) Commitments to current or future contracts for individuals having the skill and training: While an individual may have lulls in their work, allowing them to work on other projects for a defined period of time, often managers have current work or are anticipating new work that would require those individuals. Keeping those individuals, with their known skills and capabilities, on the team increases the manager's ability to execute.

2) Hiring for open positions, even replacement, is closely scrutinized: Rigorous financial targets and processes make requesting the ability to replace or hire a new individual on a team challenging for a manager. If a person, especially a highly skilled one in a non-growth area, leaves the team, then additional work is created for the manager in developing the case to replace him or her.

3) Potential candidates are the employees in demand: By virtue of having the skills and experience that identified them as strong candidates to the analytics algorithm, identified employees have very desirable skills for their current area of assignment, as well as for the growth area. Some early discussions about limiting the pool of potential candidates to those whose utilization is not high or those not meeting other targets were quickly dismissed as unsuitable for assignment to strategic areas of the business, unless they were to fit the target profile. Having a succession plan for each anticipated opening in the chain is conceptually viable, but has practical limitations.

4) External hiring is easier: By the time in the process that the hiring manager has made the case to build a team and has the financial and management agreements necessary to hire for a growth area, the external recruiting process promises quicker results and less internal negotiation. Hiring managers who did look at potential candidates identified by the process, were pleased with the results, but quickly ran into resistance from the employee's current manager, and most often were not even granted permission to have a qualification discussion with the employee. A lesson learned is that the teams less likely to find potential internal candidates already had external searches/recruiting teams in place. We should have engaged with them prior to their involvement with external hiring processes.

5) Broad transformation programs were not in place: We found ourselves in a bit of a chicken-egg dilemma: what is required is a broadly supported transformation program, but without the proof-of-concept of actually hiring and business benefits realized, an investment in a transformation program was difficult to support.

D. Other Lessons Learned

We learned several lessons in deploying the big data job transfer approach beyond the main lesson of organizational challenges to adoption. For example, we learned from examining internal job listings written by hiring managers that it is difficult for non-HR personnel less familiar with the IBM expertise taxonomy to apply it, especially for skills in emerging growth areas. On the other hand, free-form job titles are highly non-standardized and variable. Therefore, it is better to let the data of employees in the target set "speak for itself", as we have proposed, supplemented by limited additional guidance in the form of keywords.

An unexpected challenge in the process turned out to be excluding all employees already working in a growth area. Often growth areas are new or emerging areas within a company that may not have an organization structure, like department or billing codes, that can be used to identify them. Therefore, we would often inadvertently include some growth area employees in our source population. However, finding these individuals in the process helped validate the efficacy of the algorithm, especially to those who may be initially skeptical of the approach.

Another lesson is that the expertise data available to us does not capture personal factors relevant to the suitability of candidates, e.g. willingness to change work schedule or relocate. These factors can only be elicited from employees themselves once they have been highly ranked.

V. Conclusion

A. Summary

In this paper, we have developed a big data-based solution, Optigrow, for enabling the transfer of employees within a large IT services company from legacy areas to growth areas. The main data source is expertise assessment information from which we can understand the skills and competencies of employees. The data analytics algorithm creates a statistical profile for a targeted growth area team and scores employees from a broad source population across the company for suitability against that profile. We have estimated that big data-enabled internal transfers can provide very significant financial benefits to companies. We tested our algorithm with real-world IT services teams within the IBM Corporation and found the outputs to be more than satisfactory both through empirical cross-validation and through the experiences of actual hiring managers. To date however, we have not been able to facilitate actual internal transfers due to organizational barriers independent of the analytics. We remain hopeful that institutional changes will come and outline some steps in this direction below.

B. Organizational Recommendations

If one accepts the objectives of transforming the current workforce, along with the well-documented business benefits of transforming vs. "churning" by bringing new employees in while releasing other employees, then the following are recommendations for the transformation required to support an Optigrow-like process:

1) Provide incentives to the management team to participate: For the hiring managers, calibrate the hiring process to streamline hiring of candidates currently employed by the company, for example prioritizing internal searches before giving permission to hire externally. For the managers of the candidates, provide a streamlined process to back-fill selected candidates that removes the pain of negotiating for replacement resources, and also understand and respect where managers have significant short-term cost pressures.

2) Strong executive stakeholder commitment: In the management environment of a large company, the executives who are strongly committed to a transformation program may not be the executives to whom the hiring manager or a candidate's current manager report. Thus the commitment and measurements of managers on both sides of the transaction may not be sufficiently invested, or they may not have the visibility encouraging them to act. Broad executive leadership and support is therefore required, with the design point/decision making favoring movement to the growth/strategic area to help remove the organizational inertia.

3) Increase employee awareness: Employees are excited about working in strategic growth areas, but often do not realize how their skills could be applied or how they could be identified. Knowing there is a neutral vehicle such as Optigrow suggesting opportunities will increase their confidence in the organization's career opportunities.

C. Future Technical Work

In addition to the recommended organizational changes, we also recommend the following future technical research. The current algorithm does not model the similarity between different job roles, specialties, and skills or how easy it is to acquire a new skill given existing ones. If we do such analysis, either by examining historical data on skill acquisition trajectories or on the co-occurrence of skills among employees [14], we can generalize the distance metric (2) from total variation distance to, e.g., Wasserstein (earth mover's) distance.

Acknowledgment

The authors thank Yvonne Calo for leadership and vision, Jessica Fosbinder and individual hiring managers for participation and thoughtful feedback, and Aleksandra Mojsilović for support.

References

[1] H. Cao, J. Hu, C. Jiang, T. Kumar, T.-H. Li, Y. Liu, Y. Lu, S. Mahatma, A. Mojsilović, M. Sharma, M. S. Squillante, and Y. Yu, "OnTheMark: Integrated stochastic resource planning of human capital supply chains," *Interfaces*, vol. 41, no. 5, pp. 414–435, Sep.–Oct. 2011.

[2] S. Ang and S. Slaughter, "Turnover of information technology professionals: The effects of internal labor market strategies," *Data Base Adv. Inf. Sy.*, vol. 35, no. 3, pp. 11–27, Summer 2004.

[3] R. Bapna, N. Langer, A. Mehra, R. Gopal, and A. Gupta, "Human capital investments and employee performance: An analysis of IT services industry," *Manage. Sci.*, vol. 59, no. 3, pp. 641–658, Mar. 2013.

[4] A. Mojsilović and D. Connors, "Workforce analytics for the services economy," in *Handbook of Service Science*, P. P. Maglio, C. A. Kieliszewski, and J. C. Spohrer, Eds. New York, NY: Springer, 2010, pp. 437–460.

[5] J. Hu, B. K. Ray, and M. Singh, "Statistical methods for automated generation of service engagement staffing plans," *IBM J. Res. Dev.*, vol. 51, no. 3/4, pp. 281–293, May/Jul. 2007.

[6] Y. Naveh, Y. Richter, Y. Altshuler, D. L. Gresh, and D. P. Connors, "Workforce optimization: Identification and assignment of professional workers using constraint programming," *IBM J. Res. Dev.*, vol. 51, no. 3/4, pp. 263–279, May/Jul. 2007.

[7] D. R. Ilgen and J. R. Hollenbeck, "The structure of work: Job design and roles," in *Handbook of Industrial and Organizational Psychology*, M. D. Dunnette and L. M. Hough, Eds. Palo Alto, CA: Consulting Psychologists Press, 1991, pp. 165–207.

[8] K. R. Varshney, J. Wang, A. Mojsilović, D. Fang, and J. H. Bauer, "Predicting and recommending skills in the social enterprise," in *Proc. ICWSM Workshop Social Comput. Workforce 2.0*, Cambridge, MA, Jul. 2013, pp. 20–23.

[9] K. R. Varshney, V. Chenthamarakshan, S. W. Fancher, J. Wang, D. Fang, and A. Mojsilović, "Predicting employee expertise for talent management in the enterprise," in *Proc. ACM SIGKDD Conf. Knowl. Disc. Data Min.*, New York, NY, Aug. 2014, pp. 1729–1738.

[10] D. Pollard, *A User's Guide to Measure Theoretic Probability*. Cambridge, UK: Cambridge University Press, 2002.

[11] C. Scott, G. Blanchard, and G. Handy, "Classification with asymmetric label noise: Consistency and maximal denoising," in *JMLR W&CP, Conf. on Learning Theory (COLT)*, vol. 30, 2013, pp. 489–511.

[12] D. Wei and K. R. Varshney, "Robust binary hypothesis testing under contaminated likelihoods," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Brisbane, Australia, Apr. 2015, pp. 3407–3411.

[13] M. Singh, K. R. Varshney, J. Wang, A. Mojsilović, A. R. Gill, P. I. Faur, and R. Ezry, "An analytics approach for proactively combating voluntary attrition of employees," in *Proc. IEEE Int. Conf. Data Min. Workshops*, Brussels, Belgium, Dec. 2012, pp. 317–323.

[14] Y. Xu, Z. Li, A. Gupta, A. Bugdayci, and A. Bhasin, "Modeling professional similarity by mining professional career trajectories," in *Proc. ACM SIGKDD Conf. Knowl. Disc. Data Min.*, New York, NY, Aug. 2014, pp. 1945–1954.