Forecasting Career Choice for College Students Based on Campus Big Data

Min Nie[1], Lei Yang[1], Bin Ding[1], Hu Xia[1], Huachun Xu[2], and Defu Lian[1](✉)

[1] Big Data Research Center,
University of Electronic Science and Technology of China, Chengdu, China
nieminde@gmail.com, dove.ustc@gmail.com
[2] College of Teacher Education and Psychology,
Sichuan Normal University, Chengdu, China

Abstract. Career indecision is a difficult obstacle in front of adolescents. Traditional vocational assessment research measure it by means of questionnaires and diagnose the potential sources of career indecision. Based on the diagnostic outcomes, career consolers develop the treatment plans tailor to students. However, because of personal motives and the architecture of the mind, it may be difficult for students to know themselves, so that the outcome of questionnaires can not fully reflect their inner states and statuses. Self-perception theory suggest students' behavior could be used as clue for inference. Thus, we proposed a data-driven framework for forecast student career choice of graduation based on their behavior in and around the campus, playing an important role in supporting career counseling and career guiding. By evaluating on 10M behavior data of over four thousand students, we show the potential of this framework in these functionality.

Keywords: Campus big data · Career identity · Career prediction · Self-knowledge

1 Introduction

According to Erikson, the formation of a vocational identity is one of the main tasks of adolescence, and viewed as a part of the larger task of identity development [9]. Indicating the possession of a clear and stable picture of ones goal, interests and talents, vocational identity is possibly formed by sufficient career exploration and subsequent commitment at college [19]. During this period of vocational identity formation, many adolescents experience periods of indecision regarding their career [18,23]. Thus, career counseling services are essential at college, to help students make career decisions, so that special career counseling centers have even been established.

From the psychological perspective, career counseling for career indecision of college students is usually a cognitive based approach in which logical processes are employed in collecting, sifting and evaluating relevant career and personal

F. Li et al. (Eds.): APWeb 2016, Part I, LNCS 9931, pp. 359–370, 2016.
DOI: 10.1007/978-3-319-45814-4_29

information. Concretely, instruments such as Career Decision-Making Difficulties Questionnaire (CDDQ) [11] are first used to precisely diagnose the sources of students' career indecision, ranking from a lack of readiness to a lack of information about self, occupations and ways of obtaining information. Based on the diagnostic outcomes, career counselors are in a position to develop a "treatment" plan for intervening in the students career indecision.

In order to be able to engage in occupational decision making, students should first of all develop competence and skill in self-concept [24]. Because of personal motives and the architecture of the mind, it may be difficult for people to know themselves [31]. According to self-perception theory, inferring persons' internal states from their behavior is a major source of self-concept [4]. For example, if students notice that they are constantly late for class, they might rightly infer that they are not as conscientious as they thought. Since many of student behaviors are driven by internal states which are "weak, ambiguous, or uninterpretable", people can use students' behaviors as a clue to their hidden dispositions.

With the development of information technology, more and more advanced information management and monitor systems have been established in many colleges/universities, with the aims of making students life and study more convenient and efficient via smart cards. When students continuously interact within a cyber-physical space, their behaviors in and around the campus, such as having meals, shopping, borrowing books and taking courses, are accumulated in real time. These behavior data can capture different patterns that reflecting their unique habits, capability, preference and mental status, so its explosive growth in the amount has just created an opportunity for proposing a data-driven framework to help students to better know themselves.

To this end, in this paper, we propose a supervised career choice forecasting framework based on students' behavior data and career choices of graduation. Within this framework, we put forward behavior-based representative factors for affecting student career selection. These factors, supported by psychological study, include professional skills/abilities [1] learned from course-taking records, behavior order in the conscientiousness of big five personality [8], interest and preference for borrowing books, and family economic status [27] estimated by daily consumption from smart card usage. It is intuitive to cast career forecasting into a multi-class classification problem, so that algorithms such as KNN, Decision Tree or logistic regression could be used to predict his potential career choice in a determinant or probabilistic way. These multi-class classification algorithms essentially capture each college student' similarity/distance/divergence with graduates over those representative factors, rightly agreeing with the social comparison theory [10] in psychology. The central proposition of the social comparison theory is the "similarity hypothesis", which indicates that human evaluate their ability and limitations by comparing with similar others, particularly when objective and non-social means of evaluation are not available. Self-evaluation in this case probably becomes more stable and accurate. More importantly, people are especially likely to make upward comparisons, that is to evaluate themselves against successful individuals, higher probably leading to self-improvement ultimately.

In addition to supporting students to determine their career choice, in this framework, we will conduct correlation analysis between behavior-characterized factors and career choices, in order to discover the influent representative factors for affecting student career selection. Therefore, it is possible to leverage these knowledge to help students achieve their early-stated goals. For example, if we have observed the significant effect of English courses at students' career choice about going abroad for further study. Thus, students, stating this as their goal, should strive to acquire the language skills (reading, writing, listening) of English.

Finally, we evaluate the proposed framework on behavior data and career choice records of over 4,000 students. Grouping career choices of graduation into "abroad further study", "seeking jobs", "domestic further study", and "others", Micro F1-measure of the best multi-class classification algorithm could achieve 0.6 at the first semester and improve with the increase of semesters. According to the correlation analysis between behavior-based factors and career choices, we find that factors like professional skills, behavior regularity and economic status could significantly correlate with career choices.

2 Related Work

This paper forecasts career for college students based on campus big data, which could be related with both professional career mining and vocational counseling and guiding. The former one includes the prediction of expertise, skill and career movement, and recommendation of jobs while the latter one will summary the influence factors on vocational indecision and introduce some intervention techniques as well as matching theory of career selection.

2.1 Professional Career Mining

There are several directions of professional career mining, including the prediction of expertise, skills and career movement. For example, expertise can be predicted based on documents such as project descriptions, human resource databases, professional articles, program code [22], based on emails [2] due to its usage for communication about work topics, based on the use of social media [13], such as blogs, wikis, forums, microblogs, people tags and so on, and based on enterprise systems of record and data from the internal corporate social networking site as features [28]. Professional skills, being related to expertise, but having multi-label characteristics, were usually predicted by matrix completion based collaborative filtering with side information [29]. Given the volatile and unpredictable world of the work environment at present, There has been less opportunity and willingness for individuals to engage in a single organization for a lifetime [3]. Thus more and more employees begin to choose external, lateral or even downward job changes. The research topics related to career movement/mobility include uncovering a set of determinants, regularities and reproducible patterns behind career movement [7,14,30,32], ranging from movement propensity, brain circulation, vocational preference to spatio-temporal regularity and the characteristic of stratification. In addition, how to recommend tailored jobs for users when there are potential career movements is an research topics about career movement. For example, a supervised machine learning algorithm are used to recommend jobs to people based on their past job histories [20]. Due to reciprocity of job seekers and recruiters, collaborative filtering and its boost with profiles of job seekers and descriptions and requirements of jobs have been applied [16,20]. The key part in collaborative filtering is to measure (career) similarity between job seekers for either themselves or recruiters, which can be the profile/self-description similarity, career path similarity or their combination [33].

2.2 Vocational Guiding

Vocational selection is a key research topic in vocational psychology, starting with the talent matching approach developed by Parsons [21]. Matching theory was subsequently developed into the trait and factor theory of occupational choice, which necessarily measured individual talents and the attributes of particular jobs. Although it is similar to job recommendation frameworks, it placed more emphasis on individual personalities, interests, aptitudes, or other explainable and measurable characteristics by instrument tools [15]. Instead, job recommendation strives to learn the traits and factors from behavior observations, which may suffers from low interpretability. Due to its criticism in a lack of adaptiveness with the change of individual and occupational environments, developmental theories [26], career exploration theories [25], social learning/cognitive theory [17] and others was developed for determining and explaining vocational choices at different periods of life span. The goal of these theories is to help people to acquire evolving self-concept (interest, ability, motivation and need) and dynamic occupational environments, and to further make career decisions.

3 Career Choice Forecasting Based on Campus Big Data

Students' behavior in campus are continuously recorded, in cases of making payments, borrowing library books and taking courses. Forecasting career choice for students requires first to disaggregate these records into different evidence sets and then predict the career choices based on these evidence. In particular, course taking histories are used for extracting professional skills and learning mastery levels of these skills since lots of the skills equipped for future vacation is delivered by taking courses; consumption records are time-stamped so that they are leveraged for modeling the regularity of behaviors such as having breakfast and taking shower; students often borrow books for learning specific skills or expanding their knowledge, mining book borrowing preferences from their book-loan histories could benefit. Finally, since each consumption record could reflect the economic status of family [12], they are used for estimating economic status by extracting patterns, such as expenditure of each breakfast/lunch/dinner and monthly expenditure. Based on these four types of evidences, grouping career choice into four groups, i.e., "abroad further study", "seeking jobs", "domestic further study", and "others", we can leverage multi-class classification algorithm for career choice prediction.

Below, we assume that the career choices of M students $\mathcal{U} = \{u_1, \cdots, u_M\}$ and their four college years of behavior data are given. These students have borrowed N books $\mathcal{B} = \{b_1, \cdots, b_N\}$, and taken S courses $\mathcal{C} = \{c_1, \cdots, c_S\}$ in total. Each book has a category attribute, and there are T categories $\mathcal{A} = \{a_1, \cdots, a_T\}$ in total.

3.1 Learning Mastery Level of Professional Skills

As aforementioned, professional skills are extracted from course taking histories. Its each record gives students' score in the corresponding course. However, there are total thousands of courses in one university, if scores of each course are taken features, this feature representation will face with sparsity challenge. In addition, many professional skills may be determined by students' performance on several courses. For example, "Machine Learning" skills may depend on the performance on "probability and statistics", "linear algebra" and "mathematical analysis". Therefore, matrix factorization based dimension reduction algorithm will be applied for feature extractions. In particular, this algorithm takes a student-course scoring matrix as input, and maps students and courses onto the same joint latent space. Assume the scoring matrix is denoted as $R \in \mathbb{R}^{M \times S}$, whose each entry $r_{i,j}$ represents the grade of the student u_i on the course c_j, and that each student $u_i \in \mathcal{U}$ and each course $c_j \in \mathcal{C}$ are represented by points in the latent space of dimension K, denoted as a user latent factor $\mathbf{p}_i \in \mathbb{R}^K$ and a course latent factor $\mathbf{q}_j \in \mathbb{R}^K$ respectively. Each dimension of the latent space could be explained by skills according to [6], so that user latent factors represent mastery level of corresponding skills and course latent factor indicate the correlation of course with corresponding skills. The dot product between user latent factor and course latent factor approximate students' performance on courses. Student latent factors $\mathbf{P} = (\mathbf{p}_1, \cdots, \mathbf{p}_M)'$ and course latent factors $\mathbf{Q} = (\mathbf{q}_1, \cdots, \mathbf{q}_S)'$ are can be learned by optimizing the following objective functions,

$$\min_{\mathbf{P},\mathbf{Q}} \sum_{i,j} \mathrm{I}_{i,j}(r_{i,j} - \mathbf{p}_i'\mathbf{q}_j)^2 + \lambda(\sum_i \|\mathbf{p}_i\| + \sum_j \|\mathbf{q}_j\|), \tag{1}$$

where $\mathrm{I}_{i,j}$ indicates whether a student i has taken a course j or not. In other words, students' performance over those courses without being taken by students is missing.

The learning of the parameters $\mathbf{p}_i$ and $\mathbf{q}_j$ could be achieved by alternative least square or stochastic gradient descent. Alternative least square makes use of the following formula for updating parameters:

$$\mathbf{p}_i = (\lambda \mathbf{I}_K + \sum_j \mathrm{I}_{i,j} \mathbf{q}_j \mathbf{q}_j')^{-1} (\sum_j \mathrm{I}_{i,j} r_{i,j} \mathbf{q}_j)$$
$$\mathbf{q}_j = (\lambda \mathbf{I}_K + \sum_i \mathrm{I}_{i,j} \mathbf{p}_i \mathbf{p}_i')^{-1} (\sum_i \mathrm{I}_{i,j} r_{i,j} \mathbf{p}_i) \tag{2}$$

After the learning of parameters, user latent factors correspond to the features on the mastery level of professional skills.

However, when feeding the student-course score matrix into this algorithm, one important preprocessing step should be first finished. Since one course may be taught by several teachers, the grade of the same course taught by different teachers cannot probably be compared with each other due to the different teaching level. Therefore, we compute the averaging grade of each course w.r.t each teacher and then subtract it from his/her students' grade of the corresponding course. For example, a student A and a student B take a course taught by a teacher X and obtain grades r_a and r_b respectively, while a student C and student D take the same course taught by another teacher Y and obtain grades r_c and r_d respectively. The grades of the former two students are normalized as $r_a - (r_a + r_b)/2$ and $r_b - (r_a + r_b)/2$.

3.2 Modeling Behavior Regularity

Conscientious is an important personality trait and has been shown to be positively related to job/academic performance [8]. Conscientious people exhibit a tendency to show self-discipline, which just could be reflected by regularity of daily activities. Therefore, behavior regularity should be useful for help student to determine their future career choice. We particularly focus on the daily regularity of having breakfast, and going to library for the first time in each day and taking showers. Regularity of a behavior could be considered as repeatability, and will be measured by the entropy of probability that the behavior occur within specific time intervals. Assume there are n time intervals $\mathcal{T} = \{t_1, \cdots, t_n\}$, for any given student, the probability that a behavior $v \in \mathcal{V} = \{$"breakfast", "library", "shower"$\}$ will take place within time interval t_i is computed as

$$P_v(T = t_i) = \frac{n_v(t_i)}{\sum_i n_v(t_i)} \tag{3}$$

where $n_v(t_i)$ is the occur frequency of the behavior v within the time interval t_i. Then the entropy of the behavior v is computed as

$$E_v = -\sum_{i=1}^{n} P_v(T = t_i) \log P_v(T = t_i) \tag{4}$$

If the entropy of a behavior is higher, the probability over time intervals is more uniformly distributed and the regularity of this behavior is lower. When computing entropy, we assume that each time interval span half an hour with respect to all three behaviors. Since breakfast behavior is specified by time periods from

6 am to 10 am, thus the number of time intervals is 8, less than the other two cases (48 time intervals). In summary, there are three entropy features in total, to reflect students' regularity.

3.3 Mining Book Reading Interest

As introduced above, students borrow library books for learning skills outside of class to expand their knowledge. Therefore, the loan history of each student could reflect his/her interest, part of which may be correlated with future occupational choices. However, there are a great number of books which have been borrowed by students, but each student only borrow a few books among them. Thus if directly using loan frequency as evidence for borrowing preference will suffer from the sparsity challenge. One solution is to leverage dimension reduction techniques based on the loan history, similar to the extraction of professional skills. Fortunately, each book has a rich set of attributes, whose category, characterized by Chinese Library Classification[1] in this case, could be useful but more easy and more interpretable to identify the interest of students. To this end, we make statistics on the loan frequency of each category of books with respect to each student and normalize them to get a probabilistic distribution. Categories are organized as a several levels of hierarchy, but to balance between representative ability and sparse, we focus on the second level of categories in the hierarchy. Thus there are around two hundreds of categories in total.

3.4 Estimating Family Economic Status

It is possible to know family economic status by questionnaire, but students may overstate their situations to get better finical support. Therefore, it is appealing to estimate family economic status from students' consumption histories, as suggested in [12]. The consumption at different locations may play different important roles for this goal, and we put emphasis on the payment history at mess halls and supermarkets since the cost from them could occupy a very large portion of total cost. In particular, we calculate the cost of each meal and each shopping by summing multiple payments within short time interval (10 mins in this case) and compute expenditure of each day, and then convert consumption history into meal cost sequence, shopping cost sequence and daily expenditure sequence, by concatenating them by chronologically. Then we first exploit first order and second order *descriptive statistics*, including minimum, maximum, median, mean, interquartile range, standard deviation and kurtosis, as evidences for family economic status. Second, we compute the *ratio* of transaction amount per day on weekends to weekdays, and conduct Fast Four Transformation (FFT) and calculate *energy* as the sum of the squared magnitudes of each FFT component [5], which could capture consumption periodicity and thus provide another evidence for family economic status. However, in order to cancel out the effect of consumption level, the mean value of the sequence $[x_1, x_2, \cdots, x_n]$, of the length

[1] https://en.wikipedia.org/wiki/Chinese_Library_Classification.

n, should be subtracted from its each value. Based on the converted sequence $[\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_n]$ where $\tilde{x}_i = x_i - \sum x_i/n$, the energy is defined as,

$$\begin{aligned} \text{Energy} &= \frac{\sum_{i=0}^{n-1} |F_i|^2}{n} \\ F_i &= \sum_{i=0}^{n-1} e^{-j2\pi \frac{ki}{n}} \tilde{x}_i \end{aligned} \tag{5}$$

To summary, we have seven descriptive statistics, one ratio feature and one energy feature for each of three expense sequences, thus having 27 features in total.

3.5 Career Forecasting Algorithms

Due to the developmental characteristic of career identity, we should be able to predict career choice for each students in each semester. Thus, we organize these evidence by semesters. In other words, in each of the first six semesters, based on their behavior data within the semester, we extract all introduced features for each student. Then in a given semester, features of the proceeding semesters will be concatenated with the current one in an chronological order. Assume the number of features is F, the first semester includes F features, the second semester includes $2F$ features and the sixth semester includes $6F$ features. For each semester, based on the features and career choices of graduation, we train multi-class classification algorithms for career choice prediction. However, this paradigm could not enable the classifier of each semester to leverage weight of classifiers of proceeding semesters. Fortunately, these classification algorithms have relation with the additive models, such as AdaBoost. This motivates us to leverage a two-level ensemble framework for this problem. Its top level algorithm is AdaBoost with six learners, where each learner is trained on F features of each semester. Therefore, the first learner of AdaBoost is trained on features of the first semester and the last learner of AdaBoost is trained the features of the last semesters. Feature sets of different learner are disjoint with each other, we can also continue using AdaBoost algorithm for each learner. Base learners in the second AdaBoost are multi-class classification algorithms.

4 Experiments

In this section, we first introduce the dataset and the settings for evaluation. Following that, we report experimental results and deliver some discussions.

4.1 Dataset and Settings

The evaluation is conducted on a dataset from 4,246 student of the same grade. The total number of consumption records is 13,122,696, among which the records in the mess is 6,875,698. Within four years in the university, these students have borrowed 172,894 books, generating 336,238 book loan records, and taken 1,072 courses in total, generating 276,588 course-grade records. For assessing the performance of multi-class classification, we exploit widely-used Macro-F1 measure and Micro-F1 measure [34]. The former one gives equal weight to every class, regardless of its frequency, and is a per-category average of F1, and the latter one gives equal weight to every document, and is a per-document average of F1. For each of six semesters, 5-folds cross validation is performed. For the multi-class classification, we will conduct comparison between logistic regression (LR), SVM, Random Forest (RF) and Decision Tree (DT).

4.2 Results and Discussions

Comparison of Multiclass Classification Algorithms. We first show the performance comparison between four classification algorithms on the six semesters of data in Fig. 1. The performance comparison on other semesters show consistent result, and thus would not have shown here. We can see that Random Forest performs comparatively better than the others, although the difference is not large. Due to the efficiency of Random Forest, we choose it for subsequent usage.

Prediction Performance. Based on the aforementioned evaluation scheme, we study the change of prediction performance with the increase of semesters and show the results in Fig. 2. First, the results indicate that the performance of career choice prediction will improve with the increase of semesters. This may be because students will learn new skills, extent their reading interest and increase the mastering level of professional skills as time goes by. And the performance of prediction is significantly better than the Random guess (25 % by pure random and 44 % by the MostFreq class). This indicate the career choice could be

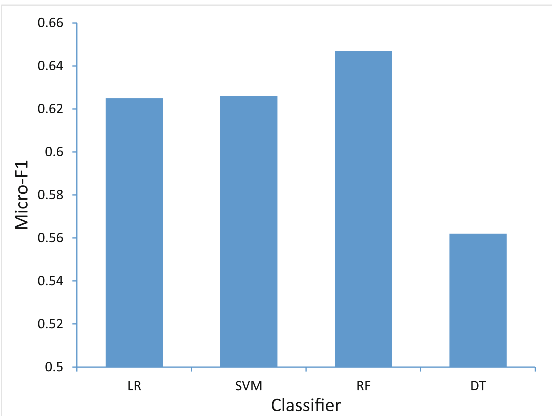

Fig. 1. The comparison of different classification algorithms

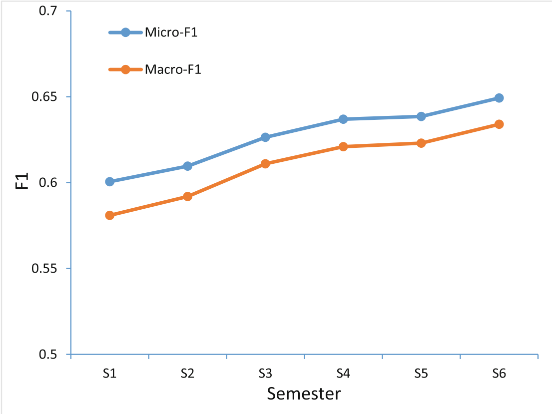

Fig. 2. Micro-F1 measure and Macro-Measure with the increasing semester

predictable from students' life and study behavior in and around the campus. In order to see the effect of different types of features, we study the feature importance for career choice prediction below.

Feature Analysis. By feed each type of features into the classification algorithm, we could obtain the performance of each type of features. The results of evaluating feature importance for career choice is shown in Fig. 3. From them, all features besides book reading interest have shown significant value for predicting career choice. And the effect of skill-based features is most salient. This should be intuitive, since one major goal that students come to university is to learn skills for preparing future occupation. Therefore, the mastery level of professional skills is directly correlated with, or determines students' career choices.

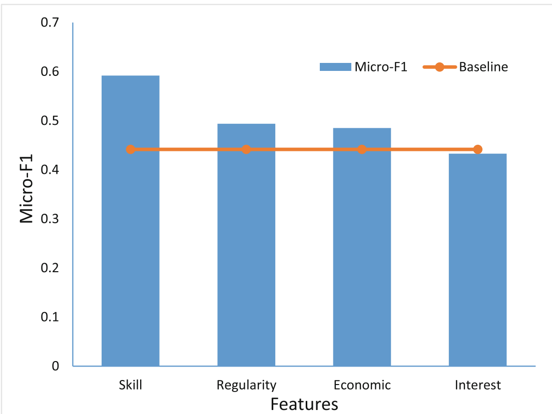

Fig. 3. Importance of four types of features by comparing them with

Behavior regularity and economic status also show significant effect for career choice prediction, confirming our aforementioned assumption. However, the reading interest reflecting in book borrowing preference is not even as well as random guess. This may lies in the sparsity of book loan history, so that it is not robust for such features to reflect students' reading interest. In the future work, it may be more appealing to leverage dimension reduction based algorithms.

5 Conclusion

In this paper, we studied career choice prediction based on students' campus behavior and proposed a data-drive framework for career choice prediction. Based on four types of behavior features, we evaluated the effectiveness of such a framework and found that the extracted professional skills, behavior regularity and economic status was significantly correlated with career choices.

In the future work, we should focus on efficient feature extraction (interest and skills) based supervised dimension reduction and further improve the performance of prediction. And we will also conduct further analysis about the relationship between student's behavior and salary.

References

1. Albion, M.J., Fogarty, G.J.: Factors influencing career decision making in adolescents and adults. J. Career Assess. **10**(1), 91–126 (2002)
2. Balog, K., de Rijke, M.: Finding experts and their details in e-mail corpora. In: Proceedings of WWW 2006, pp. 1035–1036. ACM (2006)
3. Baruch, Y.: Transforming careers: from linear to multidirectional career paths: organizational and individual perspectives. Career Dev. Int. **9**(1), 58–73 (2004)
4. Bem, D.J.: Self-perception theory (1973)
5. Chen, Y.-P., Yang, J.-Y., Liou, S.-N., Lee, G.-Y., Wang, J.-S.: Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. Appl. Math. Comput. **205**(2), 849–860 (2008)
6. Desmarais, M.C.: Mapping question items to skills with non-negative matrix factorization. ACM SIGKDD Explor. Newslett. **13**(2), 30–36 (2012)
7. Deville, P., Wang, D., Sinatra, R., Song, C., Blondel, V.D., Barabási, A.-L.: Career on the move: geography, stratification, and scientific impact. Sci. Rep. **4**, 1–7 (2014)
8. Dudley, N.M., Orvis, K.A., Lebiecki, J.E., Cortina, J.M.: A meta-analytic investigation of conscientiousness in the prediction of job performance: examining the intercorrelations and the incremental validity of narrow traits. J. Appl. Psychol. **91**(1), 40 (2006)
9. Erikson, E.H.: Identity: Youth and Crisis, vol. 7. WW Norton & Company, New York (1994)
10. Festinger, L.: A theory of social comparison processes. Hum. Relat. **7**(2), 117–140 (1954)
11. Gati, I., Krausz, M., Osipow, S.H.: A taxonomy of difficulties in career decision making. J. Couns. Psychol. **43**(4), 510 (1996)
12. Guan, C., Lu, X., Li, X., Chen, E., Zhou, W., Xiong, H.: Discovery of college students in financial hardship. In: Proceedings of ICDM 2015, pp. 141–150. IEEE (2015)

13. Guy, I., Avraham, U., Carmel, D., Ur, S., Jacovi, M., Ronen, I.: Mining expertise and interests from social media. In: Proceedings of WWW 2013, pp. 515–526. International World Wide Web Conferences Steering Committee (2013)
14. Hadiji, F., Mladenov, M., Bauckhage, C., Kersting, K.: Computer science on the move: inferring migration regularities from the web via compressed label propagation. In: Proceedings of IJCAI 2015, pp. 171–177. AAAI Press (2015)
15. Holland, J.L.: Making Vocational Choices: A Theory of Vocational Personalities and Work Environments. Psychological Assessment Resources, Odessa (1997)
16. Hong, W., Li, L., Li, T., Pan, W.: iHR: an online recruiting system for Xiamen talent service center. In: Proceedings of KDD 2013, pp. 1177–1185. ACM (2013)
17. Krumboltz, J.D., Mitchell, A.M., Jones, G.B.: A social learning theory of career selection. Couns. Psychol. **6**(1), 71–81 (1976)
18. Lopez, F.G.: A paradoxical approach to vocational indecision. Pers. Guidance J. **61**(7), 410–412 (1983)
19. Marcia, J.E., Waterman, A.S., Matteson, D.R., Archer, S.L., Orlofsky, J.L.: Ego Identity: A Handbook for Psychosocial Research. Springer Science & Business Media, New York (2012)
20. Paparrizos, I., Cambazoglu, B.B., Gionis, A.: Machine learned job recommendation. In: Proceedings of RecSys 2011, pp. 325–328. ACM (2011)
21. Parsons, F.: Choosing a Vocation. Houghton Mifflin, Boston (1909)
22. Reichling, T., Wulf, V.: Expert recommender systems in practice: evaluating semi-automatic profile generation. In: Proceedings of CHI 2009, pp. 59–68. ACM (2009)
23. Savickas, M.L.: Identity in vocational development. J. Vocat. Behav. **27**(3), 329–337 (1985)
24. Savickas, M.L.: The transition from school to work: a developmental perspective. Career Dev. Q. **47**(4), 326–336 (1999)
25. Schein, E.H.: Career Anchors: Discovering Your Real Values. University Associates San Diego, San Diego (1990)
26. Super, D.E.: A life-span, life-space approach to career development (1990)
27. Thompson, M.N., Subich, L.M.: The relation of social status to the career decision-making process. J. Vocat. Behav. **69**(2), 289–301 (2006)
28. Varshney, K.R., Chenthamarakshan, V., Fancher, S.W., Wang, J., Fang, D., Mojsilović, A.: Predicting employee expertise for talent management in the enterprise. In: Proceedings of KDD 2014, pp. 1729–1738. ACM (2014)
29. Varshney, K.R., Wang, J., Mojsilovic, A., Fang, D., Bauer, J.H.: Predicting and recommending skills in the social enterprise. In: Proceedings of AAAI ICWSM Workshop Social Computing for Workforce, vol. 2, pp. 20–23 (2013)
30. Wang, J., Zhang, Y., Posse, C., Bhasin, A.: Is it time for a career switch? In: Proceedings of WWW 2013, pp. 1377–1388. International World Wide Web Conferences Steering Committee (2013)
31. Wilson, T.D., Dunn, E.W.: Self-knowledge: its limits, value, and potential for improvement. Annu. Rev. Psychol. **55**, 493–518 (2004)
32. Xu, H., Yu, Z., Xiong, H., Guo, B., Zhu, H.: Learning career mobility and human activity patterns for job change analysis. In: Proceedings of ICDM 2015, pp. 1057–1062. IEEE (2015)
33. Xu, Y., Li, Z., Gupta, A., Bugdayci, A., Bhasin, A.: Modeling professional similarity by mining professional career trajectories. In: Proceedings of KDD 2014, pp. 1945–1954. ACM (2014)
34. Yang, Y.: An evaluation of statistical approaches to text categorization. Inf. Retrieval **1**(1–2), 69–90 (1999)