

Introduction to Quantitative Career Matching: How to Statistically Pick the Right Occupation

Cao Bittencourt

March 15, 2024

Abstract

In this introductory paper, we address the crucial, but highly neglected, subject of data-driven career choice and development. We begin by sketching out the most basic approach to quantitative career matching, and test our model by calculating career compatibility coefficients with data from the United States' primary source of occupational information, the O*NET framework. We, then, illustrate the results with a sample of occupations from O*NET's database, and describe the best matches for each selected career. Despite the model's simplicity, we found career matches to be very accurate, although not as precise as they would be had we employed superior methods. In the future, we expect to cover some advanced material, deliberately left out of this introduction, and detail more sophisticated and performant models.

Keywords: Career matching; Career compatibility; Career choice; Career development; Vocational choice; Occupational Information Network.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 5 |
| 2 | Methods | 5 |
| 2.1 | The Occupational Information Network | 5 |
| 2.2 | Skill Sets | 6 |
| 2.3 | A Euclidean Approach to Career Matching | 6 |
| 2.3.1 | An Initial Model for Euclidean Career Matching | 6 |
| 2.3.2 | A Weighted Euclidean Approach to Career Matching | 7 |
| 2.3.3 | Linear Weights for Euclidean Career Matching | 8 |
| 2.3.4 | Adjusting for Overqualification | 8 |
| 2.4 | Data and Implementation | 9 |
| 3 | Results | 11 |
| 3.1 | Most Compatible Occupations | 11 |
| 3.2 | Similarity Matrix | 14 |
| 4 | Discussion | 14 |
| 5 | Conclusion | 16 |
| | References | 17 |
| | Appendix | 18 |

List of Tables

| | | |
|---|--|----|
| 1 | General Occupational Statistics | 9 |
| 2 | Summary of Sample Occupations | 11 |
| 3 | Best Career Matches – Mechanical Engineers | 11 |
| 4 | Best Career Matches – Physicists | 12 |
| 5 | Best Career Matches – Credit Analysts | 13 |
| 6 | Best Career Matches – Dishwashers | 13 |
| 7 | Similarity Matrix | 14 |
| 8 | Detailed Skill Sets | 18 |

1 Introduction

Choosing the right occupation is one of the most important decisions in life. In fact, a considerable portion of a person’s life – generally, a third of it – is actually spent working. Additionally, finding a career that is suitable for one’s skill set and natural aptitudes greatly impacts job performance and overall satisfaction. Having sufficient compensation to feed a family, and adequate work-life balance, are also crucial. And, for some individuals, we could even cite self-fulfillment and ethical reasons as major factors contributing to the significance of work.

Given this overarching importance of career choice and development, we would expect to find many quantitative approaches to the issue in the scientific literature. Nonetheless, despite some competent efforts by psychometricians and personality theorists (e.g. Holland, 1959, 1997; Schein, 1978, 1993), not much has been done to address this problem in a purely data-driven manner.

In the late 1990s, however, a notable achievement in this area was brought about by the Occupational Information Network (National Center for O*NET Development [O*NET], 1998). Sponsored by the United States Department of Labor, this program set out to quantify job requirements in a standardized fashion. Their research resulted in a robust database of career paths, which includes numeric estimates of competence levels for dozens of abilities; as well as several other characteristics, such as the frequency of typical work activities, the seriousness of expected job hazards, etc. Their catalog of occupations is vast, and their data have been updated frequently for over two decades.

But even with such rich information publicly available, there is still a lack of statistical modeling of it. In this series of articles, we aim to fix this. Starting with the most immediate concepts, we expect to document all mathematical methods we have devised to tackle the study of career choice and development. Here, we detail a basic approach to quantitative career matching based on the estimation of compatibility scores. In the future, we shall cover more complex matching algorithms, and advanced topics, like identifying occupations’ core attributes, assessing the generality of career paths (to distinguish between generalist and specialist workers), calculating employability coefficients, constructing optimal strategies for career planning, talent acquisition, and much more.

2 Methods

2.1 The Occupational Information Network

The Occupational Information Network (O*NET) is an American primary source of occupational information. It was developed during the mid 1990s, and published in 1998, by a team of public and private sector organizations, in partnership with the Department of Labor’s Employment and Training Administration. Its aim is to facilitate the maintenance of a skilled workforce through the collection and dissemination of occupational data (Mariani, 1999; O*NET, 2024).

Essentially, the O*NET model provides a thorough, quantitative, description

of hundreds of SOC (Standard Occupational Classification, see U.S. Bureau of Labor Statistics [BLS], 2018) career paths in a vectorized format. O*NET accomplishes this task by systematically evaluating the competencies of actual people in the labor market, on a 0 to 100 scale, in order to derive numeric guidelines on what attributes are required for each position.

Therefore, the resulting “career profiles” from O*NET’s surveys are meant to be taken as “canonical representations” of SOC occupations: they “map out” what, say, an engineer, an accountant, or any other occupation tends to “look like”, in terms of professional attributes.

2.2 Skill Sets

With the convenient structure of the O*NET framework in mind, we begin by defining and measuring the professional attributes that characterize each economic agent in the labor market. To do so, we establish a bounded, uniform scale from 0 (complete incompetence) to 100 (complete mastery), such that:

$$a_i^k \in [0, 100], \quad (1)$$

where a_i^k is the i -th professional attribute of a person k . Moreover, let us denote as a “skill set”, or “career profile”, the vector of their m attributes, like so:

$$\mathbf{a}_k = (a_1^k, \dots, a_m^k). \quad (2)$$

Thus, all individuals in the economy – whether employed, or not – are mathematically described by their own skill sets, or career profiles; and this applies to regular SOC occupations as well, like the ones included in the O*NET database.

2.3 A Euclidean Approach to Career Matching

Finally, individuals and occupations are said to be similar to the exact degree to which their vectors of professional attributes, or skill sets, overlap.

To measure this compatibility, we can utilize several matching methods. Some of these are complicated, and require lengthy explanations, while others are much easier to introduce. As the scope of this article is very limited, we only cover the most basic of methods: the Euclidean approach to career matching.

2.3.1 An Initial Model for Euclidean Career Matching

In this approach to career matching, compatibility is estimated by the inverse function of Euclidean distance. Hence, we first define this concept:

$$d(\mathbf{a}_k, \mathbf{a}_q) = \|\mathbf{a}_k - \mathbf{a}_q\| = \sqrt{\sum_{i=1}^m (a_i^k - a_i^q)^2}. \quad (3)$$

Equation (3) measures the absolute distance from the professional attribute vector \mathbf{a}_k to a comparison vector \mathbf{a}_q ; in other words, it tells us how these two skill sets are *dissimilar*.

But, of course, our interest here is career compatibility, not its opposite. Furthermore, similarity is typically expressed as a percentage; and for this, we need dissimilarity to be bounded to a known interval. Therefore, in order to convert the Euclidean distance (3) to a similarity metric, we need to employ some sort of normalization procedure, like the following:

$$\tilde{d}(\mathbf{a}_k, \mathbf{a}_q) = \sqrt{\frac{\sum_{i=1}^m (a_i^k - a_i^q)^2}{\sum_{i=1}^m \max(100 - a_i^q, a_i^q)^2}}. \quad (4)$$

The denominator in equation (4) is the maximum theoretical distance to the \mathbf{a}_q comparison skill set. That is, here we normalize distance (3) by calculating the distance to \mathbf{a}_q from its most dissimilar vector. So, for each coordinate of the \mathbf{a}_q skill set, we measure which distance would be greater: that from the scale's lower bound (viz. 0), or that from the upper bound (viz. 100), as any other distance has to be less than those to the scale's limits. This way, the denominator corresponds to the maximum distance to \mathbf{a}_q ; and, consequently, the normalized distance $\tilde{d}(\mathbf{a}_k, \mathbf{a}_q) \in [0, 1] \forall \mathbf{a}_k, \mathbf{a}_q$.

At last, similarity is easily derived from the normalized dissimilarity:

$$s(\mathbf{a}_k, \mathbf{a}_q) = 1 - \tilde{d}(\mathbf{a}_k, \mathbf{a}_q). \quad (5)$$

As opposed to equations (3) and (4), this formula describes the measure to which career profiles are *similar*. Again, it is evident $s(\mathbf{a}_k, \mathbf{a}_q) \in [0, 1] \forall \mathbf{a}_k, \mathbf{a}_q$. Thus, we can estimate, as a percentage, how compatible a person is with an occupation, and whether they are alike in terms of their competencies, or not.

2.3.2 A Weighted Euclidean Approach to Career Matching

Although straightforward, the above method is far too simplistic, for each and every competency is given the same importance in matching; and this is, clearly, not reasonable. In reality, some professional attributes are, definitely, more important to some occupations, while to others they are less or even not important.

Consider, for instance, the absurdity of a barber who knows a lot about the arts, and fashion, and entertaining customers, and even accounting and bookkeeping, and yet does not know how to actually cut hair. One could call them artistic, or stylish, but hardly a barber. For despite the usefulness of all these other things for the barber, they are not nearly as important as their main activity (viz. cutting hair).

This is the issue of weighting professional competencies, and it is crucial for adequate career matching. In fact, to the degree to which attributes are said to be “central”, or “indispensable”, to certain career paths (like being able to cut hair is for a barber), so too weighting these attributes is indispensable for career matching algorithms to function properly.

Now, as it is with matching methods, here also we find a variety of manners of weighting professional attributes. The first, which we cover below, is to use linear weights. This said, quadratic, logistic, or any sort of increasing weights

can be employed as well. In our more complex career matching algorithms, for example, we make use of a specific function to determine skill sets' core competencies, and weight them accordingly, so as to vastly improve models' performance. For illustrative purposes, however, linear weights seem to suffice.

2.3.3 Linear Weights for Euclidean Career Matching

When implementing linear weights with the Euclidean approach, each distance between career profiles is multiplied by the attributes of the comparison skill set (viz. that to which compatibility is to be calculated). Thus, distances to the most important competencies are emphasized, while the remaining receive less importance.

Mathematically, the weighted metrics (with subscript w) for similarity scores and normalized distances are as follows:

$$s_w(\mathbf{a}_k, \mathbf{a}_q) = 1 - \tilde{d}_w(\mathbf{a}_k, \mathbf{a}_q), \quad (6)$$

where

$$\tilde{d}_w(\mathbf{a}_k, \mathbf{a}_q) = \sqrt{\frac{\sum_{i=1}^m a_i^q (a_i^k - a_i^q)^2}{\sum_{i=1}^m a_i^q \max(100 - a_i^q, a_i^q)^2}}. \quad (7)$$

With these adjusted equations, Euclidean distance is weighted proportionally to the professional attribute levels of the \mathbf{a}_q career profile. This means that the matching method is optimized for every comparison skill set, as irrelevant competencies are left out of the analysis, and the “indispensable” ones (e.g. cutting hair for a barber, dentistry for a dentist) are given their full importance.

2.3.4 Adjusting for Overqualification

A positive side effect of the use of weights in career matching is an implicit correction for the statistical penalties of overqualification. These arise, in the original approach, whenever an individual has additional competencies from other fields that are not required at a certain position, and cause the unadjusted distance (4) to yield less favorable similarity estimates. For, if $a_i^k > a_i^q$, then

$$\frac{\partial \tilde{d}(\mathbf{a}_k, \mathbf{a}_q)}{\partial a_i^k} = \frac{\tilde{d}(\mathbf{a}_k, \mathbf{a}_q) \times (a_i^k - a_i^q)}{\sum_{i=1}^m (a_i^k - a_i^q)^2} > 0 \iff \frac{\partial s(\mathbf{a}_k, \mathbf{a}_q)}{\partial a_i^k} < 0. \quad (8)$$

This overqualification problem is highly detrimental for matching accuracy, as it tends to “punish” people with many professional attributes, whether they are skilled generalists or merely hobbyists. But, of course, as the absurd scenario of the useless barber mentioned above, this too is not reasonable at all. Indeed, no one would say, for instance, that an airline pilot is less of a pilot if they also know how to cut hair, just as we do not say a barber is less of a barber if they cannot fly an airplane. Put another way, competencies that have “nothing to do” with an occupation should not be a limiting factor to career compatibility.

The application of matching weights helps to mitigate this, as the less important competencies are, rightly, given less importance; and those distances to completely irrelevant attributes (with competence levels of 0) are even nullified entirely, so that:

$$\tilde{d}_w(\mathbf{a}_k, \mathbf{a}_q) \leq \tilde{d}(\mathbf{a}_k, \mathbf{a}_q). \quad (9)$$

Therefore, if a person has additional, unnecessary, skills, these weighting techniques provide a correction for the penalties of overqualification, decreasing the normalized Euclidean distance and increasing compatibility.

2.4 Data and Implementation

The version of the O*NET database utilized contains 873 unique career profiles, even including a few variants on top of SOC occupations. For brevity's sake, though, they are succinctly described below in terms of general clusters:

Table 1: General Occupational Statistics

| Cluster | N ¹ | Employment ² | Wage ³ | Market Share ⁴ |
|--|----------------|-------------------------|-------------------|---------------------------|
| Business Management & Administration | 65 | 26,639,371 | \$59,924.00 | 21.05% |
| Health Science | 99 | 16,856,404 | \$62,753.00 | 13.95% |
| Marketing | 26 | 13,961,830 | \$45,512.00 | 8.38% |
| Manufacturing | 138 | 13,644,785 | \$44,169.00 | 7.95% |
| Hospitality | 42 | 17,416,290 | \$28,965.00 | 6.65% |
| Transportation, Distribution & Logistics | 62 | 11,018,200 | \$45,511.00 | 6.61% |
| Architecture & Construction | 86 | 9,005,327 | \$53,142.00 | 6.31% |
| Education & Training | 64 | 7,266,480 | \$62,621.00 | 6.00% |
| Finance | 22 | 5,255,514 | \$70,325.00 | 4.87% |
| Law, Public Safety, Corrections & Security | 35 | 4,790,620 | \$61,985.00 | 3.92% |
| Information Technology | 22 | 3,588,065 | \$82,061.00 | 3.88% |
| Human Services | 38 | 7,823,620 | \$36,809.00 | 3.80% |
| Science, Technology, Engineering & Mathematics | 70 | 2,207,815 | \$100,683.00 | 2.93% |

Continued on next page

Table 1: General Occupational Statistics (Continued)

| Cluster | N ¹ | Employment ² | Wage ³ | Market Share ⁴ |
|---|----------------|-------------------------|-------------------|---------------------------|
| Agriculture, Food & Natural Resources | 40 | 2,140,244 | \$51,134.00 | 1.44% |
| Arts, Audio/Video Technology & Communications | 40 | 1,692,270 | \$57,836.00 | 1.29% |
| Government & Public Administration | 24 | 1,166,588 | \$62,636.00 | 0.96% |

¹ Number of SOC occupations plus variants.

² Total employment levels in the United States in 2022.

³ Employment-weighted mean wages in the United States in 2022.

⁴ Total cluster wages as a percentage of total wages in the United States in 2022.

Source: BLS, 2022; O*NET, 2023.

As it concerns matching procedures, career compatibility coefficients were estimated by comparing all available occupations to one another, instead of matching against human subjects. This was done for eight reasons: to ensure the model’s internal consistency; to avoid potential sampling biases; to have more reliable “targets” when assessing accuracy (as individuals’ career profiles tend to be less defined than O*NET’s “canonical skill sets”); in a similar fashion, to make results more objectively interpretable for the reader; to simplify this initial model by postponing, for now, the making of a competence-preference distinction, that shall be needed later on to account for the fact that one’s skills are not always aligned with their preferences; because matching careers to careers is, essentially, the same as matching individuals to careers (for both are defined by competency vectors); because O*NET’s 873 occupations already provide a somewhat large sample size for this type of study; and, of course, because the O*NET data are, themselves, derived from rigorous labor market surveys, and consequently their career profiles are, in fact, representative of the skills of real people. So, even though one of the model’s end goals is to help *individuals* find the right career, at the present moment, we can test these methods “internally”, without resorting to further empirical research.

Unfortunately, this also means we cannot show most of our results, as the O*NET database contains too many career paths, and the scope of an introductory article does not allow for all their matches to be displayed. Hence, for this exercise, we only present four occupations: mechanical engineers, physicists, credit analysts, and dishwashers. These were chosen to illustrate a “sufficient range” of careers: two highly qualified STEM occupations, the first more “hands-on” than the latter; one of the simplest roles in the economy (viz. dishwashers); and the other somewhere in between (viz. credit analysts). Their main characteristics are summarized in Table 2.

Table 2: Summary of Sample Occupations

| SOC | Occupation | Cluster | Employment ² | Wage ³ |
|---------|----------------------|-------------------|-------------------------|-------------------|
| 19-2012 | Physicists | STEM ¹ | 18,840 | \$152,430.00 |
| 17-2141 | Mechanical Engineers | STEM ¹ | 277,560 | \$95,300.00 |
| 13-2041 | Credit Analysts | Finance | 71,960 | \$77,440.00 |
| 35-9021 | Dishwashers | Hospitality | 431,840 | \$28,130.00 |

¹ “STEM” stands for “Science, Technology, Engineering, and Mathematics”.

² Total employment levels in the United States in 2022.

³ Median wages in the United States in 2022.

Source: BLS, 2022; O*NET, 2023.

The career profiles used for matching were composed of three broad categories of professional attributes from the O*NET framework, namely: “Skills”, “Abilities”, and “Fields of Knowledge”. These three categories combined sum to exactly 120 competencies, and are detailed in the Appendix. The remaining categories of attributes (e.g. “Interests”, “Work Values”) do not constitute “competencies” in the strict sense of the word, so they were not included in the analysis. Additionally, a few of the original attribute names were changed to reduce ambiguity (e.g. from “Design” to “Industrial Design”).

Finally, all calculations were implemented in the R programming language, with packages we wrote and published to carry out Euclidean matching and other algorithms (Bittencourt, 2024a). The complete scripts and data can be found in an online repository (Bittencourt, 2024b).

3 Results

3.1 Most Compatible Occupations

Again, despite the small selection of four occupations, there are still too many career matches to be displayed here. Thus, we only illustrate the results of the weighted Euclidean approach with the ten most compatible occupations.

Table 3: Best Career Matches – Mechanical Engineers

| Comparison Occupation | Similarity |
|---------------------------------------|------------|
| Mechanical Engineers | 1.00 |
| Aerospace Engineers | 0.90 |
| Marine Engineers and Naval Architects | 0.88 |
| Nuclear Engineers | 0.88 |

Continued on next page

Table 3: Best Career Matches – Mechanical Engineers (Continued)

| Comparison Occupation | Similarity |
|--|------------|
| Mechatronics Engineers | 0.87 |
| Microsystems Engineers | 0.86 |
| Photonics Engineers | 0.86 |
| Fuel Cell Engineers | 0.86 |
| Automotive Engineers | 0.86 |
| Materials Engineers | 0.86 |
| Mining and Geological Engineers, Including Mining Safety Engineers | 0.86 |

Note: Similarity scores estimated with Euclidean matching and linear weights.

Source: Author's calculation based on O*NET, 2023.

Table 4: Best Career Matches – Physicists

| Comparison Occupation | Similarity |
|--|------------|
| Physicists | 1.00 |
| Astronomers | 0.85 |
| Mathematicians | 0.80 |
| Physics Teachers, Postsecondary | 0.80 |
| Materials Scientists | 0.79 |
| Engineering Teachers, Postsecondary | 0.79 |
| Nanosystems Engineers | 0.78 |
| Aerospace Engineers | 0.77 |
| Biochemists and Biophysicists | 0.77 |
| Nuclear Engineers | 0.77 |
| Computer and Information Research Scientists | 0.77 |

Note: Similarity scores estimated with Euclidean matching and linear weights.

Source: Author's calculation based on O*NET, 2023.

Table 5: Best Career Matches – Credit Analysts

| Comparison Occupation | Similarity |
|--|------------|
| Credit Analysts | 1.00 |
| Accountants and Auditors | 0.87 |
| Loan Interviewers and Clerks | 0.86 |
| Budget Analysts | 0.85 |
| Securities, Commodities, and Financial Services Sales Agents | 0.85 |
| Financial Examiners | 0.85 |
| Business Intelligence Analysts | 0.85 |
| Tax Preparers | 0.85 |
| Insurance Underwriters | 0.84 |
| Bookkeeping, Accounting, and Auditing Clerks | 0.84 |
| Tax Examiners and Collectors, and Revenue Agents | 0.83 |

Note: Similarity scores estimated with Euclidean matching and linear weights.

Source: Author’s calculation based on O*NET, 2023.

Table 6: Best Career Matches – Dishwashers

| Comparison Occupation | Similarity |
|---|------------|
| Dishwashers | 1.00 |
| Janitors and Cleaners, Except Maids and Housekeeping Cleaners | 0.88 |
| Maids and Housekeeping Cleaners | 0.88 |
| Pressers, Textile, Garment, and Related Materials Models | 0.87 |
| Postal Service Mail Sorters, Processors, and Processing Machine Operators | 0.87 |
| Sewing Machine Operators | 0.86 |
| Laundry and Dry-Cleaning Workers | 0.86 |
| Orderlies | 0.86 |
| Cutters and Trimmers, Hand | 0.86 |
| Shoe and Leather Workers and Repairers | 0.86 |

Note: Similarity scores estimated with Euclidean matching and linear weights.

Source: Author’s calculation based on O*NET, 2023.

3.2 Similarity Matrix

We also found it useful to build a similarity matrix, with which one can easily compare occupations’ compatibility coefficients:

Table 7: Similarity Matrix

| | Mechanical Engineers | Physicists | Credit Analysts | Dishwashers |
|-------------------------|-------------------------|------------|--------------------|-------------|
| Mechanical Engineers | 1.00 | 0.75 | 0.54 | 0.38 |
| Physicists | 0.80 | 1.00 | 0.57 | 0.34 |
| Credit Analysts | 0.75 | 0.66 | 1.00 | 0.54 |
| Dishwashers | 0.62 | 0.53 | 0.66 | 1.00 |

Note 1: Similarity scores estimated with Euclidean matching and linear weights.

Note 2: Comparison occupations on the horizontal axis.

Source: Author’s calculation based on O*NET, 2023.

As noted in the similarity matrix, comparison occupations are on the horizontal axis, meaning that Table 7 should be read vertically. This is because the normalization (4) and weighting (7) techniques employed imply the characteristic symmetry of Euclidean distance is, in most cases, lost.

Therefore, the table shows, for example, that mechanical engineers have 62% compatibility with dishwashers, while dishwashers have only 38% compatibility with mechanical engineers; that is to say engineers are sufficiently qualified to perform the job activities of dishwashers, but not the other way around.

4 Discussion

Assessing the matching results in Tables 3–6, we could say this Euclidean model is quite accurate: mechanical engineers are matched to ten other types of engineers; while theoretical physicists are similar to astronomers, biophysicists, mathematicians, and scientific roles, in general; credit analysts, adequately, cluster with occupations in the financial industry; and dishwashers, in turn, have high compatibility with janitors, cleaners, and so-called “blue-collar” positions.

When we analyze Table 7, we find results to be likewise accurate. Firstly, the STEM career paths are fairly close to one another, at 80% similarity. This is expected, since engineers are, in a way, applied physicists. Dishwashers, on the other hand, are not really similar to either one, but are a little more compatible with credit analysts. And, finally, credit analysts, have roughly 60% similarity with the other three occupations, indicating they are somewhat compatible, and undifferentiated in this respect. It also stands out, as already mentioned, that

the more “highly qualified” profiles have high compatibility scores with the “less qualified”, or “simpler”, ones, while the reverse is not true.

As regards precision, however, the Euclidean approach to career matching is evidently suboptimal. For one thing, similarity scores are too high, specially when comparing occupations that have little or nothing to do with one another, as we have in this exercise. That is, though the model’s matching predictions “make sense”, the degree of similarity, for some careers, is grossly overestimated.

Indeed, the compatibility of, say, dishwashers with engineers and physicists is relatively low, which is realistic; but, at almost 40%, it is still, objectively, not low enough. And the same applies to credit analysts, as well. For as an entry-level position in the field of Finance, credit analysts do seem to be some sort of “middle ground” between the other selected roles: in fact, their job activities are mostly mathematical in nature, yet are neither as demanding and abstract as those of physicists, nor as direct and concrete as the repetitive manual labor of dishwashers. So, in this sense, the estimated compatibility scores are correct. Nevertheless, the actual numeric values of these similarities are, definitely, out of proportion; and more reasonable figures would be in the range of 30% to 40%, much lower than the coefficients obtained with the Euclidean matching method.

A second, more encompassing, deficiency of this approach, that could even help to explain the first one, has to do with factors that we deliberately left out of the model. Educational attainment, for instance, is one of them, and it could drastically change matching results, as several career paths have basic requirements in the form of specific degrees, certifications, years of experience, and so on and so forth.

Of course, these other aspects are all very important, and a full analysis of career compatibility should take them into consideration. That being said, here we decided to exclude them for two main reasons. The first is that adding such variables alongside the 120 competencies without modification to the matching method does not work, in purely practical terms, because the algorithm employed is not suitable to perform binary filtering based on key, necessary conditions, but rather to compute the distance between vectors.

The second reason is that these requirements are, actually, best conceptualized as *independent* coefficients themselves, that can, then, be multiplied, or otherwise combined with any kind of function to limit career compatibility scores. Moreover, skill set similarity, the sole object of this initial model, does not imply – or is even implied by – educational attainment, years of experience, etc. In other words, as these things are separate in the real world, so too they should separate in statistical models. And this “separation of concerns” is more parsimonious and leads to a “cleaner”, theoretically consistent, framework.

At last, and most importantly, we must repeat that this Euclidean approach to career matching is an oversimplification for illustrative purposes; and that we have already developed more advanced methods to account for its problems. In fact, it is easy to derive better matching estimates only by tweaking the model presented above.

An immediate improvement, for example, would be to adjust weighting procedures to further emphasize the core competencies of comparison occupations

by substituting the linear weights with a polarizing, logistic-like function. And one could also apply scaling functions to normalized Euclidean distances, in order to correct the similarity metric directly (e.g. by means of linear interpolations, root functions, etc).

Having said this, beyond some “quick fixes”, more dramatic improvements require rewriting the entire matching algorithm to another format. But then again, the goal of this article was to provide an introduction to the crucial issue of quantitative career matching, not to exhaust the subject. In future works, we expect to be able to explain at length our sophisticated and performant models.

5 Conclusion

In this paper, we described the simplest approach to quantitative career matching, and illustrated its application with a sample of data from a robust source of occupational information, namely the O*NET framework. We also discussed a few of the fundamental issues involved in assessing career path compatibility, like competency weighting, and the problem of overqualification. In an effort to keep the article concise and accessible, however, we deliberately ignored many of the other complexities surrounding this subject.

Even so, we found our results to be mostly accurate, though not as precise as they would be had we employed superior methods. We highlight that the sample occupations were matched as “expected”: engineers with engineers; physicists with scientists; credit analysts with other roles in the financial sector; and dishwashers with janitors, and “blue-collar” positions. Nevertheless, we noted, as well, that similarity scores were too high for some of these matches.

To tackle this inconsistency, we proposed an ideal, parsimonious solution, in which a number of independent coefficients ought to be calculated to separately account for the various aspects affecting compatibility. In addition, we also proposed a couple of more immediate “quick fixes” to adjust matching estimates.

Overall, as suggested by the article’s title, this was an initial sketch to address a most important topic: that of rationally choosing a career. And, despite the basic model presented here not allowing to fully develop what “picking the right occupation” looks like statistically, we do hope and think it is, at least, a step in the right direction.

References

- Bittencourt, C. (2024a). *Atlas.match: Quantitative career matching*. <https://github.com/CaoBittencourt/atlas.match>
- Bittencourt, C. (2024b). *Scripts and data for “introduction to quantitative career matching: How to statistically pick the right occupation”*. <https://github.com/CaoBittencourt/atlas-research>
- Holland, J. L. (1959). A Theory of Vocational Choice. *Journal of Counseling Psychology*, 6(1), 35.
- Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments*. Psychological Assessment Resources.
- Mariani, M. (1999). Replace with a database: O*NET replaces the Dictionary of Occupational Titles. *Occupational Outlook Quarterly*.
- National Center for O*NET Development. (1998). *O*NET 98. O*NET Transitional Databases*. *O*NET Resource Center*. https://www.onetcenter.org/db_transitional.html
- National Center for O*NET Development. (2023). *O*NET 27.3. O*NET Database Releases Archive*. *O*NET Resource Center*. https://www.onetcenter.org/db_releases.html
- National Center for O*NET Development. (2024). *About O*NET*. *O*NET Resource Center*. <https://www.onetcenter.org/overview.html>
- Schein, E. H. (1978). *Career Dynamics: Matching Individual and Organizational Needs*. London, Addison Wesley.
- Schein, E. H. (1993). *Career Anchors – Discovering your real values*. London, Pfeiffer & Company.
- U.S. Bureau of Labor Statistics. (2018). *2018 SOC System. About SOC*. <https://www.bls.gov/soc/2018/#materials>
- U.S. Bureau of Labor Statistics. (2022). *Occupational Employment and Wage Statistics (OEWS) Survey*. <https://www.bls.gov/oes/special-requests/oesm22nat.zip>

Appendix

Table 8: Detailed Skill Sets

| Competency | Mechanical Engineers | Physicists | Credit Analysts | Dishwashers |
|-------------------------------|-------------------------|------------|--------------------|-------------|
| Engineering and Technology | 87 | 86 | 13 | 0 |
| Industrial Design | 84 | 45 | 0 | 0 |
| Applied Mathematics | 79 | 94 | 68 | 0 |
| Mechanical | 78 | 28 | 0 | 12 |
| Physics | 73 | 97 | 0 | 0 |
| Oral Comprehension | 70 | 84 | 59 | 32 |
| Reading Comprehension | 70 | 84 | 57 | 30 |
| Oral Expression | 70 | 82 | 59 | 30 |
| Deductive Reasoning | 70 | 79 | 64 | 29 |
| Pure Mathematics | 70 | 79 | 59 | 14 |
| Mathematical Reasoning | 68 | 84 | 59 | 16 |
| Written Comprehension | 68 | 79 | 59 | 29 |
| Complex Problem-Solving | 68 | 70 | 43 | 30 |
| Computers and Electronics | 67 | 83 | 43 | 0 |
| Natural Science | 66 | 82 | 5 | 0 |
| Operations Analysis | 66 | 52 | 34 | 0 |
| Active Learning | 63 | 79 | 52 | 29 |
| Information Ordering | 63 | 71 | 52 | 30 |
| Number Facility | 61 | 79 | 63 | 16 |
| Critical Thinking | 61 | 70 | 57 | 30 |
| Monitoring | 61 | 57 | 43 | 30 |
| Written Expression | 59 | 80 | 57 | 23 |
| Inductive Reasoning | 59 | 79 | 59 | 29 |

Continued on next page

Table 8: Detailed Skill Sets (Continued)

| Competency | Mechanical Engineers | Physicists | Credit Analysts | Dishwashers |
|-------------------------------|----------------------|------------|-----------------|-------------|
| Category Flexibility | 59 | 73 | 50 | 29 |
| Problem Sensitivity | 59 | 70 | 57 | 29 |
| Judgment and Decision | 59 | 64 | 54 | 30 |
| Originality | 57 | 75 | 37 | 25 |
| Writing | 57 | 71 | 55 | 25 |
| Native Language | 57 | 70 | 60 | 38 |
| Active Listening | 57 | 68 | 57 | 32 |
| Near Vision | 57 | 57 | 61 | 34 |
| Administration and Management | 57 | 54 | 42 | 37 |
| Quality Control Analysis | 57 | 46 | 11 | 29 |
| Chemistry | 56 | 61 | 0 | 21 |
| Fluency of Ideas | 55 | 73 | 41 | 27 |
| Speaking | 55 | 68 | 57 | 30 |
| Visualization | 55 | 61 | 25 | 29 |
| Systems Analysis | 55 | 57 | 43 | 16 |
| Systems Evaluation | 55 | 55 | 45 | 16 |
| Speech Recognition | 54 | 57 | 50 | 30 |
| Technology Design | 54 | 48 | 0 | 0 |
| Operations Monitoring | 54 | 34 | 14 | 30 |
| Instructing | 52 | 66 | 34 | 20 |
| Time Management | 52 | 52 | 43 | 30 |
| Troubleshooting | 52 | 29 | 0 | 29 |
| Education and Training | 51 | 72 | 29 | 41 |
| Administrative | 51 | 40 | 59 | 0 |
| Learning Strategies | 50 | 71 | 36 | 18 |
| Selective Attention | 50 | 57 | 41 | 30 |

Continued on next page

Table 8: Detailed Skill Sets (Continued)

| Competency | Mechanical Engineers | Physicists | Credit Analysts | Dishwashers |
|--------------------------------------|-------------------------|------------|--------------------|-------------|
| Persuasion | 50 | 50 | 36 | 27 |
| Coordination | 50 | 48 | 36 | 30 |
| Production and Processing | 49 | 34 | 29 | 0 |
| Speech Clarity | 48 | 71 | 45 | 27 |
| Flexibility of Closure | 48 | 55 | 45 | 29 |
| Perceptual Speed | 48 | 46 | 45 | 30 |
| Customer and Personal Service | 48 | 37 | 39 | 38 |
| Far Vision | 46 | 52 | 30 | 32 |
| Management of Personnel Resources | 45 | 45 | 27 | 29 |
| Speed of Closure | 43 | 55 | 41 | 29 |
| Negotiation | 43 | 39 | 39 | 29 |
| Social Perceptiveness | 41 | 45 | 41 | 29 |
| Service Orientation | 41 | 43 | 43 | 29 |
| Visual Color Discrimination | 41 | 41 | 21 | 29 |
| Management of Material Resources | 41 | 32 | 13 | 13 |
| Multitasking | 39 | 41 | 29 | 29 |
| Management of Financial Resources | 39 | 30 | 27 | 11 |
| Hearing Sensitivity | 39 | 30 | 16 | 29 |
| Memorization | 37 | 48 | 34 | 21 |
| Auditory Attention | 37 | 32 | 21 | 29 |
| Telecommunications | 37 | 23 | 7 | 0 |
| Equipment Selection | 37 | 21 | 0 | 29 |
| Installation | 37 | 0 | 0 | 0 |
| Building and Construction | 36 | 15 | 16 | 0 |

Continued on next page

Table 8: Detailed Skill Sets (Continued)

| Competency | Mechanical Engineers | Physicists | Credit Analysts | Dishwashers |
|----------------------------------|-------------------------|------------|--------------------|-------------|
| Personnel and Human Resources | 35 | 38 | 23 | 0 |
| Public Safety and Security | 35 | 26 | 15 | 26 |
| Programming | 34 | 55 | 14 | 0 |
| Depth Perception | 34 | 30 | 5 | 29 |
| Operation and Control | 34 | 9 | 0 | 30 |
| Sales and Marketing | 33 | 20 | 29 | 0 |
| Finger Dexterity | 32 | 29 | 25 | 34 |
| Law and Government | 32 | 28 | 54 | 28 |
| Reaction Time | 30 | 0 | 0 | 29 |
| Transportation | 28 | 12 | 0 | 33 |
| Communications and Media | 25 | 46 | 23 | 19 |
| Economics and Accounting | 25 | 26 | 74 | 0 |
| Arm-Hand Steadiness | 25 | 0 | 11 | 41 |
| Repairing | 25 | 0 | 0 | 29 |
| Equipment Maintenance | 23 | 0 | 0 | 29 |
| Medicine and Dentistry | 23 | 0 | 0 | 0 |
| Geography | 21 | 15 | 25 | 0 |
| Biology | 20 | 16 | 0 | 0 |
| Trunk Strength | 14 | 14 | 14 | 43 |
| Manual Dexterity | 14 | 0 | 9 | 52 |
| Wrist-Finger Speed | 9 | 0 | 0 | 21 |
| Rate Control | 5 | 0 | 0 | 32 |
| Spatial Orientation | 5 | 0 | 0 | 14 |
| Psychology | 0 | 28 | 10 | 0 |

Continued on next page

Table 8: Detailed Skill Sets (Continued)

| Competency | Mechanical Engineers | Physicists | Credit Analysts | Dishwashers |
|-------------------------------|-------------------------|------------|--------------------|-------------|
| Therapy and Counseling | 0 | 21 | 0 | 0 |
| Sociology and Anthropology | 0 | 0 | 16 | 0 |
| History and Archeology | 0 | 0 | 13 | 0 |
| Control Precision | 0 | 0 | 7 | 32 |
| Extent Flexibility | 0 | 0 | 0 | 43 |
| Static Strength | 0 | 0 | 0 | 39 |
| Multilimb Coordination | 0 | 0 | 0 | 37 |
| Stamina | 0 | 0 | 0 | 32 |
| Gross Body Coordination | 0 | 0 | 0 | 30 |
| Speed of Limb Movement | 0 | 0 | 0 | 30 |
| Response Orientation | 0 | 0 | 0 | 29 |
| Dynamic Strength | 0 | 0 | 0 | 25 |
| Gross Body Equilibrium | 0 | 0 | 0 | 21 |
| Glare Sensitivity | 0 | 0 | 0 | 21 |
| Sound Localization | 0 | 0 | 0 | 14 |
| Night Vision | 0 | 0 | 0 | 14 |
| Peripheral Vision | 0 | 0 | 0 | 14 |
| Explosive Strength | 0 | 0 | 0 | 11 |
| Dynamic Flexibility | 0 | 0 | 0 | 0 |
| Foreign Language | 0 | 0 | 0 | 0 |
| Philosophy and Theology | 0 | 0 | 0 | 0 |
| Fine Arts | 0 | 0 | 0 | 0 |
| Food Production | 0 | 0 | 0 | 0 |

Source: O*NET, 2023.