Measuring mismatch of skills between job vacancies and the unemployed population

Alexander Kowarik
Johannes Gussenbauer
Marlene Weinauer

The prototype, developed by Team Austria, focuses on matching skill requirements of job vacancies with skills listed in CVs by the unemployed population from the labour force. Note that the prototype was only tested on the data regarding Germany, but the calculations can be easily extended to data from any country.

Data-Sets used

In order to use as many information as possible different data sets were incorporated in the prototype. For the unemployed population the labour force survey (LFS) and the CVs from the EURES data base, and for the job vacancies the CEDEFOP data as well as the information from the job vacancy survey (JVS) were used.

Unemployed population

The CVs in the EURES data base lists ESCO-Skills as well as various demographic variables concerning the author of the CV. By using demographic variables, existing in both, the CVs and LFS, it is possible to assign skill-sets from the EURES data base to job seekers from the LFS. This matching was done using a simple algorithm, which proceeds as follows:

1. Select $d_1, ..., d_n$ demographic variables, existing in both data sets, where each variable d_i can take values from a set S_i,$i = 1, ..., n$.
2. This selection naturally splits the CVs and the LFS into disjoint subgroups CV_i and LFS_i,$i = 1, ..., \prod_{i=1}^{n} |S_i|$ defined by the values of $d_1, ..., d_n$.
3. For each subgroup LFS_i, sample with replacement $n = |LFS_i|$ skill-sets from CV_i.
4. If for an i the set CV_i is empty, discard the least significant criteria in $d_1, ..., d_n$ and calculate new subgroups. Do this only for is where $|CV_i| = 0$.
5. Repeat steps 3 and 4 until every record in the LFS-data was matched with a skill-set from the EURES data base.

Since the supplied LFS-data were public use files the matching in the prototype was done on a rather granular level. In the prototype the variables sex, ISCO Level 1 of the last occupation and age, grouped into 15-20, 25-39, 40-54, 55-74 and 75+ years of age, were used. From the EURES data base the active and inactive CVs where used as a donor set. Please note that, using this algorithm, it is easily possible to not only assign the skill-sets but also the desired occupation or desired residence to the LFS-data. This was however not done in the prototype due to time limitations.

Finally the LFS-data, inheriting the skill-sets from the EURES data base, was inflated by the supplied survey weights. This final data set is called PD for the rest of the description.

Job vacancies

The job vacancies were taken from the CEDEFOP data where for each vacant job the job-specific ESCO-skills are listed. In addition the quarterly totals from the JVS on NACE-Level were used to inflate the job vacancies on a quarterly level. This final data set is called JD for the rest of the description.

Matching Algorithm

Estimating the mismatch of skills between vacant jobs and the unemployed population is not a straight forward task, since supplied skills of the job-seekers can not directly be compared by the required skills of the vacant jobs. Instead, the supplied skills need to be treated as sets which are referenced by a person. To illustrate, consider a job market which consists of two vacant jobs J_1, J_2 and two job-seekers P_1 and P_2. The job J_1 requires skills S_1 and S_2 and J_2 requires skills S_3 and S_4. The person P_1 comes with skill-set $\{S_1, S_2, S_3, S_4\}$ and person P_2 comes with skill-set $\{S_1, S_2\}$. Directly comparing the skills supplied by the job-seekers with the skills required by the vacant jobs would seemingly indicate that there are enough skills present on the job-seeking side to occupy the vacant jobs. If however P_1 would occupy the job J_1 then P_2 would not have the necessary skills to occupy job J_2 resulting in a mismatch between vacant jobs and job-seekers. Therefore it is important to treat skills of each unemployed person as a set and consider that if that person occupies a vacant job, all of his or her skills are off the job-seeking side on the job market. It is consequently necessary for the measurement of the mismatch of skills on the job market to consider possible job occupancies by job-seeking people, and that the mismatch can, as is demonstrated by the previous example, occur by people occupying vacant jobs in such a way that other job-seekers can no longer find a suitable job. Following the same logic a job can be left vacant if job-seekers equipped with the necessary skills apply or are more suitable for different jobs, which require different skill-sets.

For our prototype we developed a sampling scheme in order to deal with the problem of estimating the mismatch of skills between vacant jobs and the unemployed population. This approach matches the unemployed population from PD to the vacant jobs from JD by comparing ESCO-Level-4-skills. In detail it works as follows:

1. For each vacant job j,$j = 1, ..., |JD|$ from JD do the following:

- Consider the job-specific ESCO-skills $s_{j_1}, ..., s_{j_m}$ for job j and define PD_j as the set of people from PD, which fulfill at least $\min(1, \lceil \alpha \cdot m \rceil)$ of those skills, with $\alpha \in [0, 1]$.
- Sample one person from PD_j by using the number of fulfilled skills per person as weights, resulting in a tuple (j, p) between a job-seeker p and a job j.
- If the set PD_j is empty flag the job j by being *unassignable*.

2. Since the sets PD_j for two different js do not need to be disjoint, a person p could have been assigned to multiple jobs in step 1. Therefore do the following:

- For each such person assign the job j for which the person fulfills the most number of skills.
- leave all the other jobs left vacant.

3. Repeat steps 1 and 2 until all jobs have an assigned person or are flagged as *unassignable*.

For the prototype α was set to 0 as we did not want to impact the results by taking an arbitrary value for α. Please note that this algorithm can easily be extended with additional Information, for example choosing PD_j through α and desired occupation or desired residence. As stated before this was not done due to time limitations. It's also straight forward to introduce additional expert knowledge in the sampling process e.g. an importance rating between job-specific skills.

Application

The application of the prototype shows a map as a front page which sole purpose is to navigate through the results of different countries. There are 3 different plots per country which show:

1. A bubble chart presenting an overview of the number of vacant jobs per quarter and job groups.
   - The size of the bubbles indicate the number of vacant jobs and the colour indicates the 'pressure', meaning the share of matched job vacancies, to the total number of vacant jobs per job group.
2. A barplot indicating the top 5 most needed skills per job group of the *unassignable* job vacancies and the impact on the job market if the unemployed population is taught each of those skills.

- The left axis on the bar plot shows the number of jobs for which this skill is job-specific and the colour coding indicates the share of job-seekers which would occupy the *unassignable* jobs if they were taught this skill.

3. A Netgraph which shows, for each job group, the most listed job-specific skills and if they often appear together on job vacancy listings.