Occupation recommendation with major programs for adolescents

Ankhtuya Ochirbat
Prof. Timothy K. Shih

Presenter: O. Ankhtuya
Outline

• Introduction
• Problem Statement
• Data Collection
• Methods
• Result
• Conclusion
Introduction

• **Young people** in high school or collage make critical decisions regarding what to study and which **career** path to pursue.

• Many of them end up switching to other majors because of **mismatching major choice** and **lack of processing information** through the professional study.
  - Such changes are **wasteful in time** and **resources** and they produce **emotional** and **economical stresses**.
• Adolescents first intend and select the **occupation** that they will work through it in the future.

• But the some **occupations** are not clear to map into the **academic program** to study or vice versa.
Why go to college?

Why go to college? Can’t I get a good job now?

Yes, maybe you could, but a college degree will make your chances even better. Check out the earnings and unemployment rates for people 25 years and older with different levels of education:

Level of education completed:
- Less than a high school diploma
- High school graduate, no college
- Some college, no degree
- Occupational program (career school)
- Associate degree (academic program)
- Bachelor’s degree
- Master’s degree
- Doctoral degree (e.g., Ph.D.)
- Professional degree (e.g., M.D., J.D.)

Unemployment rate in 2014:
- 9.0%
- 6.0%
- 6.0%
- 4.3%
- 4.6%
- 3.5%
- 2.8%
- 2.1%
- 1.9%

Mean (average) earnings in 2014:
- $30,108
- $43,056
- $48,984
- $48,152
- $52,364
- $74,308
- $88,036
- $105,456
- $124,904

Problem Statement

• But the main issue of difficulty making the decisions of major choices for the students is a lack of knowledge and information about majors and its mapping to occupations.

• Hence, it is essential to build Automatic Recommendation System.
• D. Tsogzolmaa, S.Delgermaa, and P.Ulziisaikhan (2013) investigated how to choose major properly among high school students in Mongolia.

- 27.4% of students answered choosing the major by their interest,
- 52.9% of students replied the major choice made by family member or relatives’ suggestion, and
- 15.3% of students answered that they would consider a labor market, a possibility of the finding the job as well as other factors.
• According to the M. Erdenechimeg (2013) study showed that
  ➢ 4.8% of students – vocational training center,
  ➢ 91.2% of students – college/university.

• But in the real case, 43% of high school graduates in 2013 year were enrolled in colleges or universities, the National statistical office of Mongolia reported.
Related works (3/3)

• B. Bolormaa, B. Oyunsuren, Ch. Altangerel, and Ch. Tsolmon (2015) concluded that 20.7% of 324 students have not yet chosen the major; the reason is that students have a lack of occupational information a lot, and schools have not provided the guidance and counseling.

• Hence, they recommend that Online automate system, which can provide occupational information including contexts, requirements, tasks and competence needed for the particular majors/occupations and the guidance/counseling for students are needed.
What is Recommender techniques?

Recommender techniques are information agents that attempt to predict which items out of large pool a user may be interested in and recommend the best ones to a target user.

Examples:

- Amazon
- IMDb
- Netflix
- LinkedIn
- Google
- Yahoo!
- YouTube
- Facebook
- The New York Times
Recommender Systems

Inputs

- Spatial
- Temporal
- Personal
- Demographic
- Psychographic
- Behavioral

Preprocessing

- Distance Measures
- Sampling
- Dimensionality Reduction
  - PCA
  - SVD

Analysis

Classifying, Clustering

Output

Predictions, Recommendations, Patterns
Person-Environment-Occupation Performance (PEOP) model

www.LifelongLearningWithOT.wordpress.com

**Physiological**
- Sleep, nutrition, strength, flexibility, inactivity, stress

**Cognitive**
- Process of thinking: memory, reasoning, attention

**Spiritual**
- Beyond religion: what has meaning for the person?

**Neuro-behavioural**
- Systems that control motor & sensory inputs eg balance, coordination

**Psychological**
- Personality, self-esteem, mental health, self-awareness, motivation

**Social Support**
- Practical or emotional support from interpersonal relationships

**Social & Economic systems**
- Political or economic policies affecting housing, health, employment

**Culture & Values**
- Customs, beliefs, traditions of a group or society

**Built Environment & Technology**
- Buildings, public spaces, tools

**Natural Environment**
- Natural terrain, climate

**Structure of tasks**
- OCCUPATION
  - OCCUPATIONAL PERFORMANCE AND PARTICIPATION
  - Well-Being
  - Quality of Life

**References:**
Case study: PEOP model & Occupational choice
Occupation Recommender System (ORS) is to assist adolescents in discovering the occupation interests.
Methods

• **Recommender Techniques category:**

- Collaborative filtering (CF)
- Content-based approach
- Hybrid models (CF + Content-based approach)
- Neighborhood-based approach
- Model-based approach
- Hybrid models (Neighborhood-based + Model-based approach)
Collaborative Filtering (CF)

• CF techniques use the datasets of preferences for items by users to predict additional topics or products the active user might like.

• CF consist of three main categories:
  - memory-based,
  - model-based, and
  - hybrid CF algorithms
Collaborative Filtering

• **Memory-based Collaborative Filtering**
  - It utilizes the entire user(student)-item(occupation) dataset to generate a prediction.
  - Uses user-item ratings matrix
  - Makes item-to-item correlations (similarity)
  - Finds items that are highly correlated
  - Recommends items with highest correlation
Data normalization

- I converted those scores into **fuzzy values** in range [0,1] that can be representing in graphical view which is a membership function using Trapezoid function.

\[
\mu_{O \mu_{S}}(x) = \begin{cases} 
1 & x < a \\
\frac{x - a}{b - a} & a \ll x \ll b \\
\frac{d - x}{d - c} & c \ll x \ll d \\
1 & x > d
\end{cases} \quad (1)
\]

**Figure 3.** Membership function for interests of students and occupations
Similarity computation

- **Fuzzy similarity** measures a similarity between item $i$ and $j$ is given by

\[
\text{sim}(i, j) = \frac{\sum_{k=1}^{m} \min(x_{ik}, x_{jk})}{\sum_{k=1}^{m} \max(x_{ik}, x_{jk})}
\]  

(6)

Where $j = 1, \ldots, n$

$n =$ total numbers of occupation

$m =$ number of factors in vocational interest

$i =$ is the current student

$x_i, x_j =$ Scores of student and occupation in interest
Prediction/Recommendation computation

- **Prediction**: weighted sum method computes the it.
  - Each rating is weighted by the corresponding similarity \( S_{i,N} \) between items \( i \) and \( N \). \( R_{u,N} \) is the rating of user \( u \) on item \( N \). Prediction \( P_{u,i} \) as

  \[
  P_{u,i} = \frac{\sum \text{all similar items } N (S_{i,N} * R_{u,N})}{\sum \text{all similar items } N (|S_{i,N}|)}
  \]

- **Top-N recommendation:**

```plaintext
Algorithm 1: Top n occupation recommendation algorithm

Inputs: studentID, occupations, n, factors
Output: Top-n recommended occupations for studentID

1: for every occupation i derived from questionnaire for studentID
2:   for every occupation j in database
3:     compute a similarity ss between i and j
4:     add studentID's preference for j, weighted by ss, to a running average
5: return the top-n occupations, ranked by weighted average
```
Model-based CF

- Model-based CF algorithms involve building a model based on the dataset of ratings.

  Regression model.
Regression model

• The model is developed using regression analysis to find patterns based on training data.

• Vocational interests (six factors - RIASEC) are predicted by using personality and learning style of students in (10) - (15).

\[
R = 0.47 \times openness + 0.06 \times sensory + 0.07 \times visual + 1.02 \tag{10}
\]

\[
I = 0.61 \times openness + 0.05 \times visual + 1.17 \tag{11}
\]

\[
A = 0.78 \times openness + 0.06 \times visual + 0.66 \tag{12}
\]

\[
S = 0.34 \times Extraversion + 0.19 \times Agreeableness + 0.14 \times Consciousness + 0.27 \times neuroticism + 0.36 \times openness + 0.07 \times active - 1.01 \tag{13}
\]

\[
E = 0.35 \times Extraversion + 0.20 \times neuroticism + 0.47 \times openness + 0.04 \times active + 0.03 \times visual - 0.44 \tag{14}
\]

\[
C = 0.35 \times Extraversion + 0.23 \times neuroticism + 0.42 \times openness + 0.06 \times active + 0.04 \times visual - 0.38 \tag{15}
\]
Variables associated with adolescents include:

- Demographic Information;
- Personality;
- Interests;
- Learning style; and
- Intended major;
Data Collection - Person

• **Participants**
  - 190 students were participated in the study.
    - This study was carried out during the first semester of the 2014/2016 academic year in Mongolia.
    - 81 in grade 10, 107 in grade 11, and 2 in grade 12.
    - The participants’ age range is from 15 to 19 years.
    - Female = 107 and male = 80
Data Collection - Person

- Person - Big Five Inventory (BFI - Goldberg, 1992, 1993)

- The BFI has the 44 questions with responses made on a Likert scale format ranging from 1 = strongly disagree to 5 = strongly agree.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraversion</td>
<td>- Friendliness, Gregariousness, Assertiveness, Activity Level, Excitement-Seeking, Cheerfulness</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>- Trust, Morality, Altruism, Cooperation, Modesty, Sympathy</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>- Self-efficacy, Orderliness, Dutifulness, Achievement-striving, Self-discipline, Cautiousness</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>- Anxiety, Anger, Depression, Self-consciousness, Immoderation, Vulnerability</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>- Imagination, Artistic Interests, Emotionality, Adventurousness, Intellect, Liberalism</td>
</tr>
</tbody>
</table>
Data Collection - Person

- **Index of Learning style**
  
  (ILS - Soloman & Felder, 1999)

  - People take in and process information in **different ways** based on their individual preferences.
  - ILS contains 44 **two choices questions**.
  - 4 **dimensions**
Data Collection - Person

- **Vocational Interest:**
  - Holland code model
  - *(RIASEC - Holland, 1985a, b, 1997; Holland & Gottfredson, 1992)*
  - It has the **106-item** with responses made on a **5-point scale** (strongly dislike to strongly like). There are **six types**.
  - All occupations can be classified according to their alignment with these six types.
Data Collection - Occupation

- Career cluster
- Education
- Task
- Technology and Tools
- Alternate title (DOT)
- Academic program (CIP)
- Knowledge
- Ability
- Interest
- Skill

Relationships:
- Occupation has Knowledge
- Occupation has Ability
- Occupation has Interest
- Occupation has Skill
- Occupation classifies to Career cluster
- Education requires level of
- Task performed by
- Occupation uses
How to connect and measure the correlation between student and occupation?

- **Vocational Interest (Holland code model)** RIASEC - 6 factors

<table>
<thead>
<tr>
<th>Occupation Description</th>
<th>Range: 1-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chief Executives</td>
<td>1.33 2.00 2.67 3.67 7.00 5.33</td>
</tr>
<tr>
<td>Chief Sustainability Officers</td>
<td>1.00 4.33 2.67 2.33 7.00 4.33</td>
</tr>
<tr>
<td>General and Operations Managers</td>
<td>1.33 1.33 1.00 3.33 7.00 3.67</td>
</tr>
<tr>
<td>Legislators</td>
<td>1.00 3.67 3.67 4.67 7.00 3.00</td>
</tr>
<tr>
<td>Advertising and Promotions Managers</td>
<td>1.67 2.00 5.33 2.33 7.00 4.67</td>
</tr>
<tr>
<td>Green Marketers</td>
<td>1.00 5.33 4.33 2.33 5.33 3.00</td>
</tr>
<tr>
<td>Marketing Managers</td>
<td>1.00 2.33 3.67 2.67 7.00 5.33</td>
</tr>
<tr>
<td>Sales Managers</td>
<td>3.00 2.00 2.00 3.67 7.00 4.67</td>
</tr>
<tr>
<td>Public Relations and Fundraising Managers</td>
<td>1.33 1.33 5.00 3.67 7.00 3.67</td>
</tr>
<tr>
<td>Administrative Services Managers</td>
<td>2.00 2.33 1.00 2.67 7.00 5.33</td>
</tr>
<tr>
<td>Computer and Information Systems Managers</td>
<td>4.00 4.33 1.67 1.67 6.67 5.33</td>
</tr>
<tr>
<td>Treasurers and Controllers</td>
<td>1.67 2.67 1.00 2.67 6.00 7.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student Interest from Questionnaire</th>
<th>Range: 1-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>studentID</td>
<td>r</td>
</tr>
<tr>
<td>669</td>
<td>3.74</td>
</tr>
<tr>
<td>673</td>
<td>3.32</td>
</tr>
<tr>
<td>677</td>
<td>3.37</td>
</tr>
<tr>
<td>681</td>
<td>3.95</td>
</tr>
<tr>
<td>685</td>
<td>3.53</td>
</tr>
<tr>
<td>689</td>
<td>2.68</td>
</tr>
<tr>
<td>693</td>
<td>2.74</td>
</tr>
<tr>
<td>697</td>
<td>3.16</td>
</tr>
<tr>
<td>701</td>
<td>4.53</td>
</tr>
<tr>
<td>705</td>
<td>3.63</td>
</tr>
<tr>
<td>709</td>
<td>3.79</td>
</tr>
<tr>
<td>713</td>
<td>3.53</td>
</tr>
<tr>
<td>717</td>
<td>3.37</td>
</tr>
<tr>
<td>721</td>
<td>2.95</td>
</tr>
</tbody>
</table>
Figure 17 Frequencies of students on Vocational interests, Big five personalities, and Learning styles
Table 9 Pearson’s correlations between suggested, intended, predicted, and recommended majors

<table>
<thead>
<tr>
<th>Majors</th>
<th>1.</th>
<th>2.</th>
<th>3.</th>
<th>4.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Suggested major by parents</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Intended major of students</td>
<td>.362*</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Predicted major</td>
<td>-.020</td>
<td>.164*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4. Recommended major by Holland model</td>
<td>.086</td>
<td>.141</td>
<td>.228*</td>
<td>1</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.
Figure 19 Correlations between high score, favorite subject and four kinds of majors. Dashed lines indicate the correlations. (*p < .05, **p < .01)
### Table 10 Influence of the choice of the major

<table>
<thead>
<tr>
<th>Questions of the influence of the major choice of the student</th>
<th>Yes</th>
<th>No</th>
<th>Not filled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Will you discuss with your parents when you make a critical decision?</td>
<td>133 (70%)</td>
<td>20 (10.5%)</td>
<td>37 (19.5%)</td>
</tr>
<tr>
<td>Will you follow the suggestion of your parent, if your parent asks you to choose their suggested major, even you don’t like?</td>
<td>30 (15.5%)</td>
<td>121 (64%)</td>
<td>39 (20.5%)</td>
</tr>
<tr>
<td>Why did you choose this intended major?</td>
<td></td>
<td></td>
<td>37 (19.4%)</td>
</tr>
<tr>
<td>According to my interest</td>
<td>130</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easily finding a job</td>
<td></td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Family (specially parent) suggested</td>
<td></td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Nowadays this major is very popular</td>
<td></td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>High salary</td>
<td></td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Teacher suggested</td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>My friends are going to choose</td>
<td></td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
### Result (5/5)

<table>
<thead>
<tr>
<th>Method</th>
<th>R</th>
<th>I</th>
<th>A</th>
<th>S</th>
<th>E</th>
<th>C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression Model</td>
<td>0.22</td>
<td>0.12</td>
<td>0.06</td>
<td>0.08</td>
<td>0.13</td>
<td>0.11</td>
<td>0.1215</td>
</tr>
<tr>
<td>HC Model</td>
<td>0.17</td>
<td>0.13</td>
<td>0.11</td>
<td>0.09</td>
<td>0.14</td>
<td>0.12</td>
<td>0.1261</td>
</tr>
</tbody>
</table>

**Table 5:** Influence of the choice of the major. Lower value is better result. RMSE is used.
Future Works

• Open Government Data

• MOOC

• Open Educational Resource
  - Wikipedia

• Linked Open Data
Thank you!