New York University Shanghai

Reading for Session 10: Integer Programming

BUSF-SHU 210: Business Analytics (Spring 2019)

In Session 10, we are going to learn another prescriptive analytics method: Integer programming. In some applications, the decision variables of an optimization model must take integer values. Such models are called integer programs. In this session, we will be discussing an exciting application of integer programming in online dating: eHarmony.

eHarmony

As discussed in the first session of this course, eHarmony is an online dating website that takes a scientific approach to love and marriage. eHarmony focuses on long term relationships, and has become a huge success in this market. Indicators of success for eHarmony include the fact that nearly 4% of U.S. marriages in 2012 are a result of eHarmony.

To help its users establish long-term compatible relationships, eHarmony uses a compatibility score between two people and uses optimization algorithms to determine the user's best matches. More specifically, eHarmony calculates a compatibility score between two people, and applies optimization algorithms to determine the user's best matches. The compatibility score is based on 29 different dimensions of personality including character, emotions, values, traits, etc. The calculation of the score involves a questionnaire of 436 questions covering these areas. Figure 1 is a screen shot of one question in eHarmony's questionnaire. A match between a man and a woman might be formed if they are compatible in 25 out of the 29 personality areas.

A key difference between eHarmony and other online dating platforms is that, rather than letting their users browse each others' profiles, eHarmony takes a quantitative optimization approach to match-making. Specifically, based on the similarity of users' answers to the questionnaire and historical data, eHarmony predicts a **compatibility score** between each pair of man and woman. The compatibility score prediction is based on the predictive analytics methodologies we have learned in the first half of this course, such as logistic regression, clustering, and trees. In this reading (and Session 10), we focus on, given the predicted compatibility scores between different pairs, how we can use prescriptive analytics approaches to optimize the matching of the users.

A Specific Case

Let's now consider a specific matching problem for eHarmony. Suppose we have three men (M_1, M_2, and M_3) and three women (W_1, W_2, and W_3). Their compatibility scores range between 1 and 5 for all pairs, as shown in Figure 2. For example, the first man and the first woman have a compatibility score of 1. The first man and the second woman have a compatibility score of 3, and so forth. We summarize the compatibility scores of all 9 possible pairs in Table 1.

eHarmony needs to form a matching so that each man is exactly matched with a woman, and each

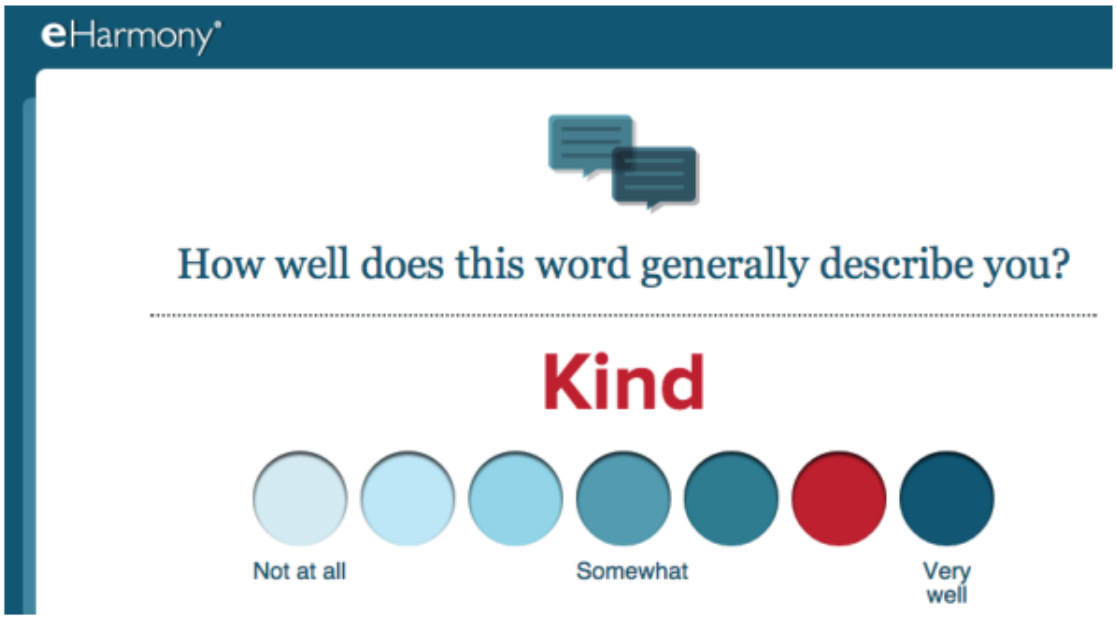

Figure 1: Questionnaire of eHarmony

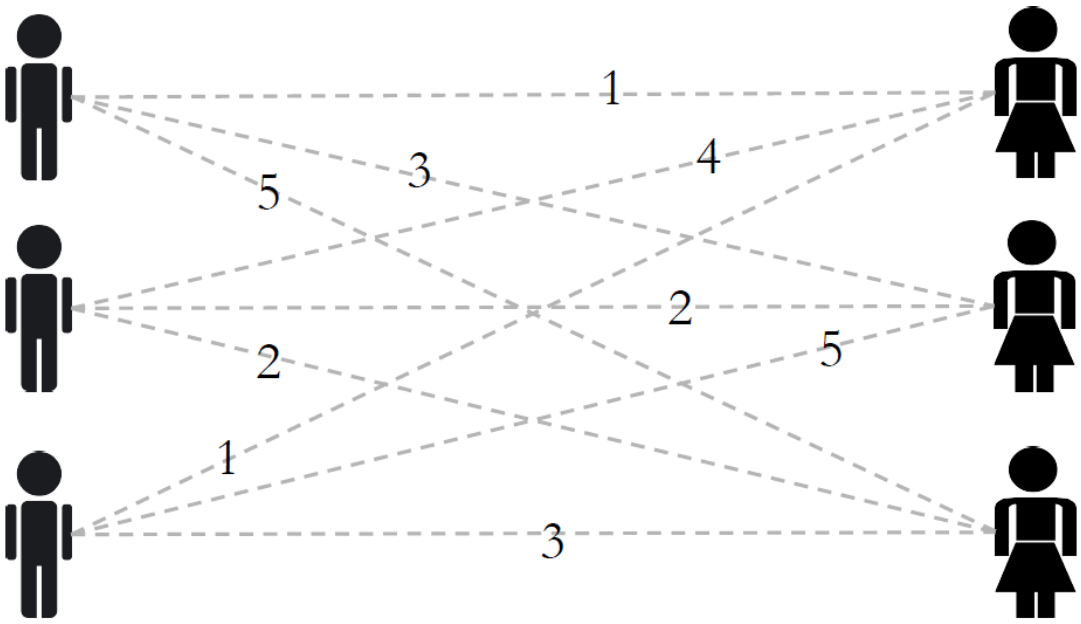

Figure 2: Compatibility Scores

	W_1	W_2	W_3
M_1	1	3	5
M_2	4	2	2
M_3	1	5	3

Table 1: Compatibility Scores

woman is exactly matched with a man. For example, in Figure 3, the red lines indicate a feasible matching. In this case, M_1 is matched to W_3 with a compatibility score of 5. The M_2 is matched to W_1 with a compatibility score of 4. Finally, the M_3 is matched to W_2 with a compatibility score of 5. Total compatibility score of this match is thus 5+4+5=14. eHarmony's goal is to find a feasible matching that **maximizes the total compatibility score**.

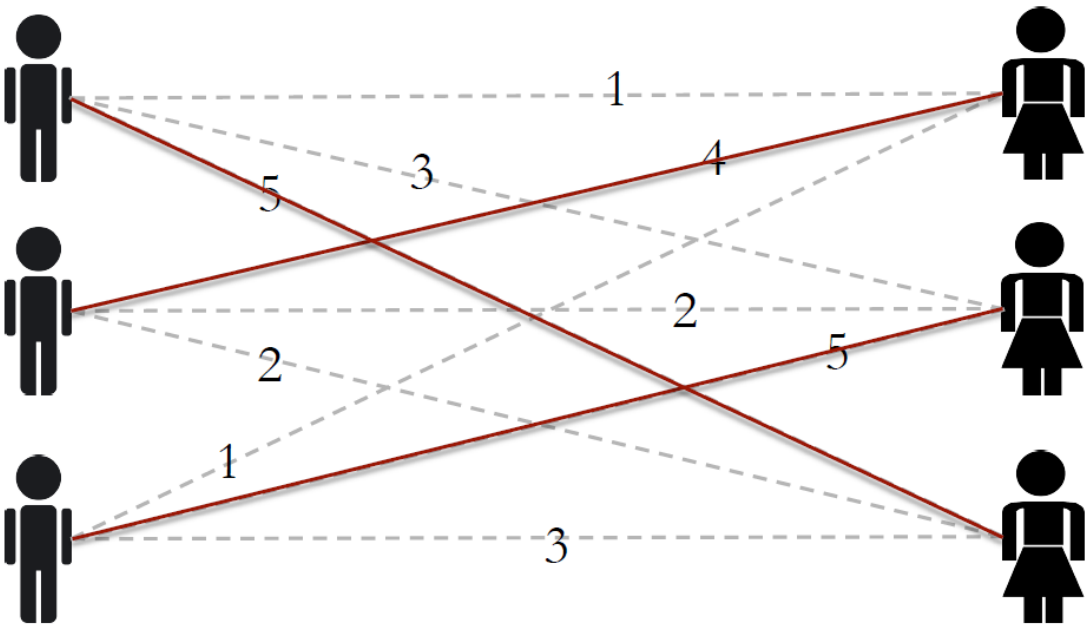

Figure 3: A Feasible Match

Questions

Now we try to formulate the matching problem described above as an optimization model. Let $x_{ij} \in \{0, 1\}$ $(i, j = 1, 2, 3)$ be the decision variables. For each i and j, $x_{ij} = 1$ if we match M_i to W_j; and $x_{ij} = 0$ if we do not match M_i to W_j. For the matching illustrated in Figure 3, $x_{13} = x_{21} = x_{32} = 1$, and $x_{11} = x_{12} = x_{22} = x_{23} = x_{31} = x_{33} = 0$.

To prepare for Session 10, try to answer the following questions before the class starts. Consider

the problem described above.

(a) What is the objective function for the matching problem?

(b) What are the constraints for the matching problem?

(c) Can you find an optimal matching that will yield the maximum total compatibility score?

We are going to discuss the case of eHarmony in Session 10. Please answer the questions above before the class starts, but no need to submit your solutions.