In the optimization problem, we assumed the compatibility scores were data that we could input directly into the optimization model.

But where do these scores come from?

In the words of the founder-- Neil Clark Warren --opposites attract, then they attack.

eHarmony's compatibility match score is based on similarity between users' answers to the questionnaire.

Let us attempt to demonstrate an approach to develop compatibility scores.

We utilize public data from eHarmony containing features for 275,000 users and binary compatibility.

Feature names and exact values are masked to protect users' privacy.

Correspondingly we won't be able to directly interpret which features are important as we do not know the identity of these features.

We used logistic regression on pairs of users' differences to predict compatibility.

To reduce the size of the problem, we filtered the data to include only users in the Boston area who have compatibility scores listed in the data set.

We computed absolute difference in features for these 1,475 pairs and trained a logistic regression model on these differences.

Let us observe the results of this experiment.

If we use a low threshold in the logistic regression model, we predict more false positives but also get more true positives.

For example, the classification matrix for threshold equal to 0.2 is as follows.

Note that we found 1,030 pairs that are not compatible and 92 pairs that are compatible correctly.

Note that 92 out of 319-- which is 227 plus 92 --of these were correctly identified.

That is, 29% percent of the matches we recommend would be successful, a very high success rate for online dating.

Clearly, there is a potential for using many other analytic methods.

Specifically trees, which are especially useful for predicting compatibility if there are nonlinear relationships between variables.

Clustering is another potential approach with the idea of segmenting the users.

Finally, text analytics is yet another approach with the idea of analyzing the text of users' profiles.

Of course, many other techniques are possible.

To give some intuition of various features, let us see how the probability of a match changes with the distance between the two adults.

It is interesting to note that the probability drops with distance, and then for a very long distance, the probability increases again.

Also interesting is this graph that shows that if the attractiveness is too high or too low, the probability of a successful match decreases.

Finally, if the difference in height is too high or too low, the probability of the match also drops.

It seems the sweet spot is a difference in height between four and eight inches.