Uncovering the Truth of Love: A Modern Approach

Kyle (Chentao) Wang
cwang556@wisc.edu

Feiyu Yue
fyue3@wisc.edu

Tz-Ruei Liu
tliu292@wisc.edu

1. Introduction

Throughout the long history of human beings, love has been a super popular, exciting, and long-lasting topic in daily life. Love is said to be a kind of magic or miracle. Statistically, it actually is if you consider it in this way: given a person, try to find another one who he or she can call as his or her "soul mate" from more than 7.7 billion people living on Earth (excluding other creatures and potential aliens to reduce the "heavy" computing followed) [1]. The probability of finding such a unique match could be as low as 1.2987013e-10 (if we consider this as it follows a Binomial Distribution). Love is also a mystery as it cannot be simply defined or explained. People have been trying to find what is the thing that connects two people with a variety of different "features" together closely. There are at least thousands of poems or books talking about what love is by Literature people [2]. Neuroscience or Psychology people may define love as a property of a complex mixture of neuropeptides and neurotransmitters [3]. As Statistic, Mathematics, or Computer Science(s) people, even though we are always considered as indifferent to romance, we still find models (in mathematics instead of those in real-life) to be sexy enough to help with defining love.

2. Motivation

According to a study from The Washington Post, approximately 51 percent of Americans between the ages of 18 and 34 said that they do not have a steady romantic partner. In fact, this 2018 figure has increased significantly from 33 percent in 2004 [4]. However, it's not just the United States that is experiencing this high single proportion. Recently, the United Nations gathered data for roughly 100 countries, showing how marriage rates changed from 1970 to 2005. Marriage rates indeed "had fallen in fourth-fifths of the countries" [5]. Ford Torney, a 26-year-old man in Baltimore, USA, really wants a steady partner but he just cannot find the right connection despite wandering through numerous date apps. He does not know how to find a partner that matches him the best and he needs to keep reminding himself that most people are not married within the country, and he is only an outlier within his own social group [4].

Figure 1. Speed dating [6]

Many people are experiencing the same thing as Torney and have been struggling to find the person that best matches his or her interest. Therefore, our group is interested in finding out how well two people are matched using machine learning tools learned in class. Given one individual's features including gender, position, country, income, and etc, our group aims to find out whether another individual is a perfect match for him/her. By applying the machine learning algorithms including decision trees and random forests, we would like to help people to overcome the partner barrier by providing everyone the best choice.

In addition, if there are some features that appeal to people the most, we would like to explore them more and illustrate how these features can increase the probability of an individual being successfully matched. These will be regarded as valuable resources for people struggling to find a romantic partner–to become someone that attracts others the most.

3. Evaluation

We hope to train models that can predict whether two people would like to see their date again after a fast "first date", which serves as a pre-survey. In other words, given two people's information with their first-date's feedback, our final model is expected to tell whether they will have a second date (return 1) or not (return 0) with high level of accuracy.

Here we use two measurements as follows to evaluate our model.

1) Accuracy: $accuracy = \frac{n_{correct}}{n_{total}}$
2) AUC: the area under the ROC curve consisting of true positive rate (TPR) on x-axis against the false positive rate (FPR) on y-axis

All the evaluations will be based on the test set. We consider our model to be successful if the accuracy is over 70%, besides, since accuracy will partly depend on the threshold we set, to balance such effect, the AUC of our model should also be greater than 0.7.

Regarding the real world application, our model could serve as decision aids as well as recommend system applied to the social networking software, online dating websites and etc.

4. Resources

4.1. Dataset

Our dataset is provided by Kaggle dataset: Speed Dating Experiment[7], it was first was compiled by professors Ray Fisman and Sheena Iyengar from Columbia Business School for their paper *Gender Differences in Mate Selection: Evidence From a Speed Dating Experiment*[8].

4.2. Approach

We specify this problem as a decision problem. We are going to use decision tree as our initial benchmark although the detailed choice is still to be discussed. Then we plan to improve our model by random forest which, as an ensemble learning method, constructs a multitude of decision trees and is expected to have a better performance on avoiding overfitting. However, random forest algorithm tends to focus more on those attributes with variant values and generally has less accuracy then gradient boosting decision tree (GBDT) algorithm. Hence we consider to use XGBoost as our final model, which is an efficient implementation of gradient boosting with the regularization on the loss function.

4.3. Computation Environment

All algorithms will be implemented in Python using Jupyter Notebook. In particular, our model training will be supported by the fundamental packages such as NumPy and other libraries like Matplotlib, Scikit-learn.

5. Contributions

Tz-Ruei Liu plans to do data cleaning and data visualization. Kyle Wang will be mainly in charge of preparing the presentation including the speech and the slides. Feiyu Yue will be responsible of evaluating and testing models. All three members will explore the body of work, including feature engineering, model implementation, ensemble learning as well as optimization. We will also divide the report evenly among group members.

References

[1] "Current world population." [Online]. Available: https://www.worldometers.info/world-population/

[2] M. Popova, "What is love? famous definitions from 400 years of literary history," Mar 2016. [Online]. Available: https://www.brainpickings.org/2013/01/01/what-is-love/

[3] K. Seshadri, "The neuroendocrinology of love," *Indian Journal of Endocrinology and Metabolism*, vol. 20, no. 4, p. 558, 2016.

[4] Bonos and Guskin, "It's not just you: New data shows more than half of young people in america don't have a romantic partner," *The Washington Post*, 2019.

[5] Zagorsky, "Why are few people getting married?" *The Conversation*, 2019.

[6] A. Shen, "The secret to getting the second date," Aug 2018. [Online]. Available: https://www.kaggle.com/aeshen/the-secret-to-getting-the-second-date

[7] A. Montoya, "Speed dating experiment," 2016, (Accessed: 16 October 2019). [Online]. Available: https://www.kaggle.com/annavictoria/speed-dating-experiment

[8] R. Fisman, S. S. Iyengar, E. Kamenica, and I. Simonson, "Gender differences in mate selection: Evidence from a speed dating experiment," *The Quarterly Journal of Economics*, vol. 121, no. 4, pp. 673–697, May 2006.