Decision-Making in Romantic Relationships

Data Modeling to Predict and Analyze Psychological Decisions in Posed Romantic Contexts

IEOR 166: Decision Analytics

Team 9

Anita Tam, Majd Muna, Carol Lee, Noor Gill

I. Introduction

While industries such as finance and marketing depend on concrete analysis with tech-powered tools to optimize their decisions, this is not the case for decisions within romantic relationships. Despite technological advancements, it is difficult to quantify aspects of a relationship between humans, and individuals depend on an ancient method for decision-making in their romantic relationships — their own psychological processes. We reference similar studies and utilize tools from the course such as utility theory and probability models to develop a logit model aimed to effectively capture one's priorities in a romantic context to predict their decision in said contexts.

Anita Tam performed data cleaning and normalization, feature selection, and model results analysis. Majd Muna contributed to data analysis, model cross-validation, feature selection, implications, delimitations, and research. Carol Lee managed data balancing, modeling, and methodology. Noor Gill performed exploratory data analysis and data visualizations, model objectives, purpose, delimitations, and future plans. The team all managed survey formulation and administration, final report writing, and presentation slides.

II. Purpose and Motivation

When faced with tough decisions involving relationships, people usually weigh their options both logically and emotionally before coming to a decision they believe best fits their priorities and values. We're familiar with the notion of complex, mathematical models used to predict the changing market or future product sales, and we are highly interested in exploring if we could do something similar with something highly qualitative and difficult to map: the psychological decision-making process in romantic contexts. The gravity of relationship decisions resonates plentiful with many personally, which fuels our intent to better understand and connect with others psychologically.

III. Objectives and Goals

We aim to answer through our study:

1. What factors drive decision-making in relationships? How does this change as the situation becomes more complex?
2. Is it possible to illustrate an individual's values in certain romantic contexts using a numerical model and produce an accurate and "effective" predictor of their response in these situations?
3. To what extent should this model be incorporated into our romantic lives?

IV. Methodology

We aim to answer our questions through a case study of human decision-making in situations common in romantic relationships. Specifically, we will focus on the maintenance of long-distance relationships, communication in those relationships, and decisions regarding marriage. Our case study consists of three main steps: first, perform a large-scale survey to understand what decisions an individual will make given a romantic context and a set of priorities. Since priorities are subjective for each individual, the survey collects data on individual rankings of priorities in addition to the decision that one makes in response to the posed context. Second, train a logit model using a subset of the data as a training set. The ranked priorities in each scenario serve as the independent predictor variables, with strong variables chosen as features in the final model. Finally, analyze the results. We apply the model on the test set and analyze the predictions generated to gain insight into the effectiveness and accuracy of our model in capturing the psychological decision-making process mathematically.

Large-Scale Survey

We presented three scenarios, each one building off of the previous in context and complexity. We presented each one in the following format: 1) provide the decision context, 2) have the individual assign values to a set of given preferences, 3) provide potential outcomes, and 4) have the individual choose a decision from a given list. The first scenario describes a long-distance relationship and asks the individual to rank several priorities then choose the decision to commit to the relationship or to break up. The second scenario adds on an additional layer of complexity of deciding whether to pause communication with your partner, break up, or suggest a compromise in response to your partner's needs, with additional priorities. The third scenario adds on the decision of marriage, with again more priorities and different decisions to make. The full scenarios with their descriptions, priorities, and decisions can be found in the survey in the appendix.

We successfully administered this survey through a Google form to 84 individuals, most of whom were in the age range of 18-24 years old. Individuals had ranked each priority on a scale from 0 to 10, with 0 being not important and 10 being extremely important. To effectively compare rankings between individuals, we normalized the data so that the sum of each individual's rankings equaled 100. We assume that individual preferences follow the von Neumann–Morgenstern axioms of orderability, transitivity, and independence, meaning individuals can order their priorities that follow the VNM axioms and priorities are assumed to be independent of each other.

Model Training

We decided on using logistic regression as a decision predictor to understand how one's likelihood for choosing an alternative changes as their values for different priority features change. The logit model specifies the probability that an individual chooses an option using a linear combination of an individual's values for each priority, which resembles one's utility function. For each scenario, we first performed a train/test split on the data, with 80% for training and 20% set aside for testing so that we could assess the model performance on unseen data. We randomly split the data while ensuring that the proportion of each class was the same in both sets. Next, we noticed that our data had heavy class imbalances, with one decision alternative having many more data points than the other for all scenarios. We tried out several methods of data balancing, including adjusting the weights of the data points and oversampling to create new data points for the minority class. Finally, we used 5-fold cross-validation to prevent overfitting to the training data and to select the best features for our model.

Final Models

For scenario 1, we found that Random Oversampling performed the best for data balancing. This method randomly samples points from the minority class to become additional data points until it is the same size as the majority class. For scenarios 2 and 3, we used the Synthetic Minority Oversampling Technique (SMOTE) to draw a line between the examples in the feature space and randomly select a new sample at a point along that line, creating new samples for the minority classes until the classes were all the same size.

After data balancing and cross-validation, we finalized our models to run on the test set and analyze the results. The priority features used in the final model for each scenario is seen under the 'Priorities/Features' columns in the tables under the Results section.

V. Results

Survey Response Results

For scenario 1, we observe that 91.8% of our respondents chose to continue committing to their relationship rather than breaking up. In scenario 2, 87.1% of respondents chose to lessen or temporarily pause communication with their partner rather than breaking up. In scenario 3, 67% of respondents chose to hint or suggest marriage, 28.2% chose to dismiss family's and friends' concern about marriage, and 4.7% of respondents chose to break up.

Model Parameters and Analysis

Scenario 1:

We denote the following:

π_0 : The probability an individual chose to break-up with their partner

π_1 : The probability an individual chose to continue to commit to their relationship

$\beta_{1,i}$:model parameter for priority i, $i \neq 0$

$I_{s,j}$: model intercept for alternative j under scenario s

v_i: the value/ranking an individual places on priority i, ranging from 0 to 10

We aim to model the log-odds of an individual choosing alternative 1 over alternative 0 as a linear combination of the feature variables in the model for scenario 1.

$$log\left(\frac{\pi_1}{\pi_0}\right) = \beta_{1,AGE} \cdot age + \beta_{1,COMMITMENT} \cdot v_{COMMITMENT} + \beta_{1,DISTANCE} \cdot v_{DISTANCE} + \beta_{1,CAREER} \cdot v_{CAREER} + \beta_{1,PARTNER\ CAREER} \cdot v_{PARTNER\ CAREER} + I_{1,1}$$

Our model gives us the following values:

$$log\,(\frac{\pi_1}{\pi_0}) \;=\; 0.152\,\cdot\,age\;+\;0.033\,\cdot v_{COMMITMENT}\;-\;0.08\;\cdot v_{DISTANCE}$$
$$+\;0.07\,\cdot v_{CAREER}\;-\;0.15\,\cdot v_{PARTNER\,CAREER}\;-\;0.001$$

Since scenario 1 poses a question that focuses on a long distance relationship partially due to one's partner's career, we expect the priorities/features distance and partner's career to stand out in weight. As expected, we see that age, the value an individual places on proximity to their partner and the value placed on their partner's career seem to be the main driving factors in one's decision to scenario 1. For every unit increase in one's value of distance to their partner, the log odds of them choosing to commit to their relationship over breaking up decreases by 0.08. For every unit increase in one's value of their partner's career, the log odds of them choosing to commit to their relationship over breaking up decreases by 0.15. At this stage, it seems like modeling one's priorities in a romantic context as a function of utility to be used in the logit model seems to capture the real-life psychological decision-making process to maximize one's goal in a relationship fairly well.

On the test data set, our final model produced a 66.7% accuracy rate in predicting people's choices for scenario 1. Accuracy for predicting class 0 individuals is 100% and that for class 1 is 62.5%.

Scenario 2:
In addition to the situation in Scenario 1, your partner is now suggesting to pause communication because of work purposes.

We denote the following:

π_0 : The probability an individual accepts their partner's proposal to temporarily pause communication

π_1 : The probability an individual suggests to lessen communication

π_2 : The probability an individual suggests to break-up with their partner

where $\beta_{1,i}$, $I_{s,j}$, and v_i are the same as in Scenario 1.

Since we have 3 alternatives in this scenario, our log likelihood for each alternatives becomes:

$$log\,(\frac{\pi_i}{\pi_i^C}) \;=\; \beta_{2,\,i,\,AGE}\cdot\,age\;+\beta_{2,\,i,\,COMMITMENT}\cdot v_{COMMITMENT}\;+\;\beta_{2,\,i,\,DISTANCE}\cdot v_{DISTANCE}$$
$$+\;\beta_{2,\,i,\,COMMUNICATION}\cdot v_{COMMUNICATION}\;+\;\beta_{2,\,i,\,PARTNER\,TIME}\cdot v_{PARTNER\,TIME}$$
$$+\;\beta_{2,\,i,\,,SECURITY}\cdot v_{SECURITY}\;+\;\beta_{2,\,i,\,,COMPROMISE}\cdot v_{COMPROMISE}+I_{2,\,i}$$

for alternative *i, i = 0, 1, 2*

Table 1 summarizes our model parameters for all 3 alternatives:

Table 1: Scenario 2 Log-Likelihood Weights For Each Priority

Features / Priorities	Temporarily Pause Communication ($\beta_{2,0}$)	Suggest Communicating Less ($\beta_{2,1}$)	Suggest to Break-up ($\beta_{2,2}$)
Age	0.317	0.388	-0.706
Commitment to the Relationship	0.0217	0.139	-0.161
Physical proximity / distance to partner	-0.052	-0.112	0.164
Communication	- 0.420	-0.305	0.725
Your partner's time	0.51	0.568	-1.078
Your sense of security in the relationship	-0.058	-0.099	0.157
Your will to compromise with your partner	-0.003	-0.191	0.194
Intercept, $I_{2,i}$	- 4.518	- 6.188	10.706

Since this scenario is focused on the topic of communication and time in addition to long distance and career concerns, we expect the priorities 'communication' and 'time' to stand out. Here, we see that age, communication and partner's time to show the most drastic differences in weights across the different alternatives.

Our model says that if we keep all other variables constant, for every one year increase in age, the log likelihood an individual will choose to break up compared to the other two alternatives decrease by 0.706. Similarly, for every one unit increase in one's ranking for 'communication,' the log likelihood they will choose to break up compared to the other two alternatives increases by 0.725. For every one unit increase in one's ranking for their "partner's time," the log likelihood they will choose to break up compared to the other two alternatives decreases by -1.078.

Again, it seems our model captures intuition fairly well. It's expected that one who values communication in a relationship would be less likely to commit to one with scarce or no communication. Similarly, one who values their partner's time would be more likely to agree to less or no communication as that aligns with their values most.

On the test data set, our final model produced a 55.6% accuracy rate in predicting people's choices for scenario 2. Accuracy for predicting class 0 individuals is 40%, class 1 is 60%, and class 2 is 66.7%.

Scenario 3:
In addition to the situation in Scenario 1 and 2, the respondent is now facing concern from family and friends about marriage.

We denote the following:

π_0 : The probability an individuals suggest marriage to their partner

π_1 : The probability an individual dismisses family/friend suggestions of marriage

π_2 : The probability an individual hints at marriage to their partner

π_3 : The probability an individual suggests to break-up with their partner

Our log-odds model for scenario 3 is:

$$log\left(\frac{\pi_i}{\pi_i^c}\right) = \beta_{3,i,AGE} \cdot age + \beta_{3,i,COMMUNICATION} \cdot v_{COMMUNICATION} + \beta_{3,i,TIME} \cdot v_{TIME} + \beta_{3,i,PARTNER\,TIME} \cdot v_{PARTNER\,TIME} + + \beta_{3,i,FAMILY-FRIENDS} \cdot v_{FAMILY-FRIENDS} + \beta_{3,i,,MARRIAGE} \cdot v_{MARRIAGE} + I_{3,i}$$

for alternative i, $i = 0, 1, 2, 3$

Table 2 summarizes our model parameters for all 4 alternatives:

Table 2: Scenario 3 Log-Likelihood Weights For Each Priority

Features / Priorities	Suggest marriage to your partner ($\beta_{3,0}$)	Dismiss family/friend suggestions of marriage ($\beta_{3,1}$)	Hint at Marriage to Your Partner ($\beta_{3,2}$)	Suggest to Break-up & Move on ($\beta_{3,3}$)
Age	0.073	-0.098	0.105	-0.08
Communication	-0.054	-0.194	-0.148	0.395
Your time	-0.734	-0.079	-0.263	1.076
Your partner's time	0.566	0.330	0.343	-1.240
Your family's/friends' satisfaction	0.119	0.288	0.108	-0.516
Your desire to get married	0.103	-0.346	-0.042	0.284
Intercept	2.604	4.896	1.754	-9.254

The features that stand out here are one's time, their partner's time, their family/friends' satisfaction, and their own desire for marriage. The weights for one's time and their partner's time fits matches intuition in that the more one values their own time, the more likely they are to move on in the posed situation and the more one values their partner's time, the more likely they will move towards marriage or temporarily dismiss family/friends' concerns.

However, we do see mismatch between expectation and our model weights for family/friends' satisfaction and one's desire for marriage. Though we would expect that the more one values their family/friends' concerns, the less likely they are to dismiss those concerns, this is opposite of what the model is saying. Likely, the model believes that the more one values their own desire for marriage, the more likely they will choose the alternatives involving marriage, though our model is telling us that the log likelihood of one wanting to break up increases the most.
On the test data set, our final model produced a 44.4% accuracy rate in predicting people's choices for scenario 2. Accuracy for predicting class 0 individuals is 75%, class 1 is 60%, and class 2 is 25%, and class 3 is 0%.

VI. Implications

To contextualize the final model's fundamentals and capabilities, we studied three pillars:

- **Effectiveness**: We recognized that our model nicely identifies the typical priorities that go into making a relationship decision. We did that via feature selection after seeing each respective priority's correlation with final decisions. The model also has decent accuracy in predictions given the simplicity of features used.
- **Usability**: Although the model is successful in mechanizing the human-human relationship aspect in romance, we understand that our model is best utilized as a reference for decision-making rather than a final ultimatum. This is due to the lack of survey data which do not encompass all decision classes thoroughly, impacting the confidence in model accuracy at this point in time. Further model complexity can potentially make the model more usable and more accurate.
- **Qualitative Understanding**: A key philosophical and societal takeaway from our findings is that society tends to choose alternatives that allow them to generally maintain their relationship, even with additional concerns and decision complexities. This is concluded via the majority favoring alternatives that maintain relationships in all scenarios of the survey.

VII. Delimitations

Possible delimitations of the project include the small size of the sample and the consequent narrow age ranges of survey respondents. The sample size has 84 individuals which means that we lacked data to have a more accurate representation of our population. The majority, if not all, of these individuals are also within the 18-24 age range (college students and young adults). This group isn't representative of the entire population, but rather a same generation with similar perspectives or maturity levels in terms of relationships.

Another delimitation is the generalization of relationship situations, such as limited scenarios and standardized intimacy levels which do not fully represent relationship ebbs and flows. We had 3 increasingly complex relationship scenarios which do not fully reflect the vast dynamics of relationships or have the capacity to test individual priorities thoroughly. Additionally, the surveys standardize the time spent in the relationship and also do not consider intimacy levels, which could act as confounders in decision-making.

VIII. Conclusion

Summary

Over 3 months of research and analysis, we sought out to learn if it is possible to create a mathematical model that predicts relationship-related decisions in certain scenarios based on individual values as utilities. To do that, we surveyed 84 people in 3 increasingly complex relationship scenarios, analyzed how their priority rankings affect their final decisions, and used those rankings as utility measures which followed the von Neumann-Morgenstern axioms. After using a logit model to predict each individual's decisions, our results showed that model accuracy typically decreased as the complexity of scenarios increased. However, the model captured many psychological intuitions in relationship decision-making. Despite sample and generalization limitations, modeling romantic decisions this way generally yields accurate results and may be a good tool to aid romantic decision-making upon further exploration.

What We Learned

Our experience with this project was quite enriching and helped us further understand the connection between algorithms and relationship decision-making. For instance, we learned that rational relationship decision models need personalized predictions for each unique individual, rather than standardized ones. Interestingly, we found out that rankings of relationship priorities change depending on the situation, and algorithms need to account for that. Intuitively, we identified that more priorities get factored into decision-making when situations grow more complex. Perhaps most importantly, we achieved our objective by concluding that models can be built based on an individual's priorities to give personalized predictions and direction towards rational decision-making.

Future Applications

Looking forward to future applications, we can improve the model through data collection by collecting survey responses from a larger and more diverse sample with age boundaries outside of the 18-24 year range we focused on. We can perform more detailed feature engineering to determine the optimal features to account for in our model and include a wider range of scenarios to test all possible values. In terms of the model, we can experiment with decision trees which would be relevant given the context of our project, try out the probit model, and also tune hyperparameters after further research. Finally, with a hybrid approach, we can consult with relationship experts or therapists to view their perspectives and discuss the potential for a hybrid tech model.

Acknowledgements
We would like to thank Professor Righter, Isaac Meza Lopez, and our peers for inspiring us, directing us, and giving us the opportunity to fulfill our project goals. It was an enriching experience!

IX. Resources and References

Barron, F. Hutton, Detlof von Winterfeldt, and Gregory W. Fischer. "Empirical and theoretical relationships between value and utility functions." *Acta Psychologica* 56, no. 1-3 (1984): 233-244

Bierlaire, Michel. *Discrete choice models*, pp. 203–227, Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.

Hoffman, Saul D. and Greg J. Duncan, *Multinomial and conditional logit discrete-choice models in demography*, Demography 25 (1988), no. 3, 415–427.

Kornblum, Angelika, Dana Unger, and Gudela Grote. "How romantic relationships affect individual career goal attainment: A transactive goal dynamics perspective." *Journal of Vocational Behavior* 125 (2021): 103523.

Logg, Jennifer M., Julia A. Minson, and Don A. Moore. "Algorithm appreciation: People prefer algorithmic to human judgment." *Organizational Behavior and Human Decision Processes* 151 (2019): 90-103.

McFadden, Daniel et al., *Conditional logit analysis of qualitative choice behavior*.

Spanier, Graham B., and ERIK E. Filsinger. "Dyadic adjustment scale." *Handbook of measurement for marriage and family therapy* (1987): 52-58.

Theil, Henri. "A multinomial extension of the linear logit model." *International economic review* 10, no. 3 (1969): 251-259.

Yoshida, Elise M. "Understanding how deciding and relationship confidence predict relationship satisfaction." (2013).

X. Appendix

Survey Form
Python Code
Cross-Validation Code
Presentation Slides