How to be confident and passionate in a romantic relationship

20170859 Dang Minh Hieu
20170866 Hoang Duc Kien

I. Introduction

Love is one of the most important aspects of our life. After decades of research, we have known much more about love and the brain, but we only understand a little about how to improve our love life [1]. Moreover, data tells us that relationships are still failing at an alarming rate [2], which have serious negative impacts on each individual life, including psychological or emotional problems, financial problems, legal problems, and problems with family [8]. In this work, we focus on one aspect of relationship: How to be confident and passionate. We will focus mainly on demographic and behavioral data, and point out how much each behavioral or demographic factor affects confidence and passion.

II. Related Work

There are many deep studies on factors contributing to love[2][3]; however, rare studies focus on what factors should be improved for each individual [1]. We come across several works to measure love satisfaction[3], which contributes to the longevity of relationships. One study from Illinois State University [2] classifies interactions in a relationship into categories: maintenance behavior, conflict behavior (constructive and destructive), relationship characteristics, and also abuse, to quantify satisfaction. Agishtein et al[7] inspects love from the cultural and diversity perspectives, and claims that the difference in social circumstances can improve or reduce the quality of the relationship, contributing to its endurance, but the study lacks individuals' differences in other aspects. Some surveys proposed by Inga et al[3], and Fehr et al[5] include personalities similarity, but judging one's own personality is too biased for any survey takers. Joel et al [6] use data from multiple spectra, which is a reliable study for feature extractions. The factors are analyzed using statistical tests, or ANOVA methods [2], and some Machine Learning models are applied: Linear Regression [3] and Random Forest [6]. Famous for its simplicity and interpretability, Linear Regression lacks the flexibility to achieve high accuracy, compared to Random Forest, which is not easy for data inference. Finally, those studies lack individual assessments for each user, which is essential that people want to know what they do correct or incorrect in their relationship.

III. Problem Definition

We focus on 2 main goals in this research

a. How to make people confident about their relationship?

To understand how couples should change in order to improve confidence in a relationship. We will try to point out which factors (e.g. how old each person is, are they texting frequently, etc.) have a strong relationship with their confidence.

b. How to make people passionate about their relationships?

To understand how couples should change in order to improve confidence in a relationship. We will try to point out which factors (e.g. how old each person is, are they texting frequently, etc.) have a strong relationship with their passion in a relationship.

IV. Methodology

a. Data collection

We mentioned 2 different sources of data in the proposal, but due to technical issues (we cannot get enough users and it is very hard to collect data on a large scale), we will only collect and work on the survey data from Mturk.

We launched a paid survey, where we pay responders some small incentive upon finishing our survey. In total, there were 362 responders, however, we only approved 202 users after filtering out noisy answers.

Survey details
We asked the following question in the survey:

1. You are currently in a romantic relationship?
2. You are currently in a long-distance relationship?
3. How old are you?
4. How old is your partner?
5. What is your ethnicity?
6. What is your partner's ethnicity?
7. Your income is (above, about, or below average)?
8. Your partner's income is (above, about, or below average)?
9. How many years have you been to school?
10. How many years has your partner been to school?
11. How many years have you been working in your current job?

12. How many years have your partner been working in your current job?
13. From 0 - 10, how much do you love your partner?
14. From 0 - 10, how often do you text your partner?
15. From 0 - 10, how often do you call your partner?
16. From 0 - 10, how often do you two talk about the future?
17. Do you two call each other by a nickname?
18. From 0 - 10, how often do you spend time WITHOUT your partner-?
19. From 0 - 10, how often do you celebrate a special day together (birthday, graduation ceremony, 1 year together, etc.)?
20. From 0 - 10, how often do you and your partner do leisure activities with your partner (e.g. cooking together, cycling together)?
21. From 0 - 10, how confident do you feel that your relationship will last for at least 1 more year?
22. Are you using any application for couples (e.g. Luveri?)
23. You are NOT in a relationship?

We filter the responses in the following way

- If the answer to question 1 and question 23 are contracting, we reject the response
- If “years of work” + “years in school” + 5 (when they are too young to either work or study) > “how old are you”, we reject the response
- If “years of work of partner” + “years in school of partner” + 5 (when they are too young to either work or study) > “how old are you of partner”, we reject the response
- If the responder finishes the survey in less than 100 seconds, we reject the response (since the average time people finish our survey is 327 seconds, and mean - 3 * standard deviation is 127 seconds)
- Otherwise, we accept the response

By this filter, we remove all the answers that are (~99%) random responses. The accepted answers, even though there are still some noises, are reasonably clean. We checked their distributions and the data’s distribution seems natural.

The data is available here (including noise data, which are marked in the feedback column): https://drive.google.com/file/d/15lIDxRJhNdELz181APqn4Ifdy-msx7f9/view?usp=sharing

b. Modeling Methods

From the survey, we set the features as multiple choices, or score bars for users’ convenience, making our features categorical rather than continuous. Thus, it is reasonable to utilize Categorical distribution statistics. To compare the difference in behaviors of users based on their belief in the future, as well as their passion in love , we implement methods regarding distribution comparison for each feature:

- Statistical test for distribution comparison: Chi-square test [9] has been widely applied to judging distribution difference between samples

- Information Value [10] is based on Information theory and is frequently used by FinTech companies for Logistics models. Different thresholds for this quantity also helps us to compare the strength between different features.
- Correlations [11] is well-known for its easy-to-interpret linear relationship. Since two statistics above grant only importance of the feature, we want to check correlations to see how the features affect our target

In addition, we will use a simple Linear Regression [12] to fit in the important variables, to quantify how they will affect the users' relationships. Due to the goal of exploration rather than prediction, we do not split the data into sets for the model.
Full Implementations can be found here:
https://drive.google.com/file/d/1ZAUHFOYCXHLZajQOaIP7HHtwgscPSnWp/view?usp=sharing

c. Ethical concern

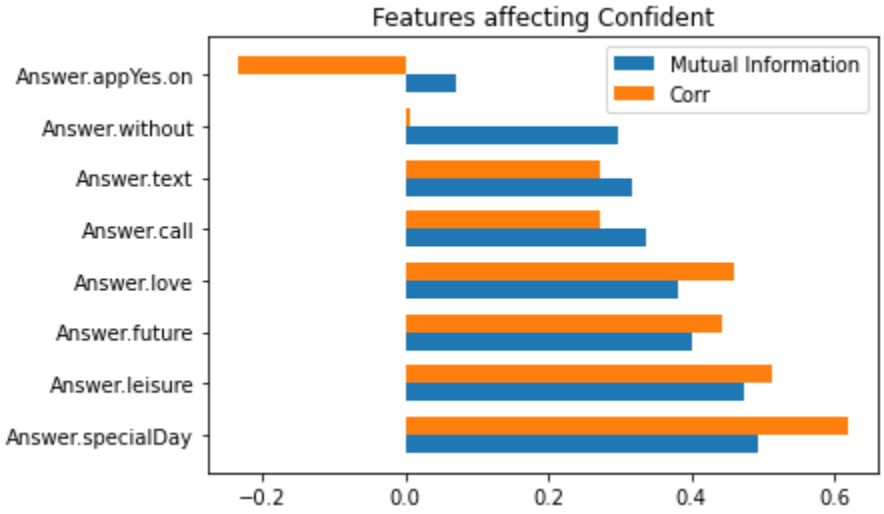
We used Mturk, an application that pays anonymous users to do our tests, so it is safe for us to use and explore the data freely. Moreover, due to the anonymity, we can publish our dataset, which is contained in the link above, without any harm.

V. Result

a. Features Exploration

For features exploration with regard to our targets, we will utilize all 3 statistics above for features' importance. If 2 out of 3 statistics show a significance in a relationship, we will conclude that the feature is important in the users' romance.

Graph for features affecting Confidence:



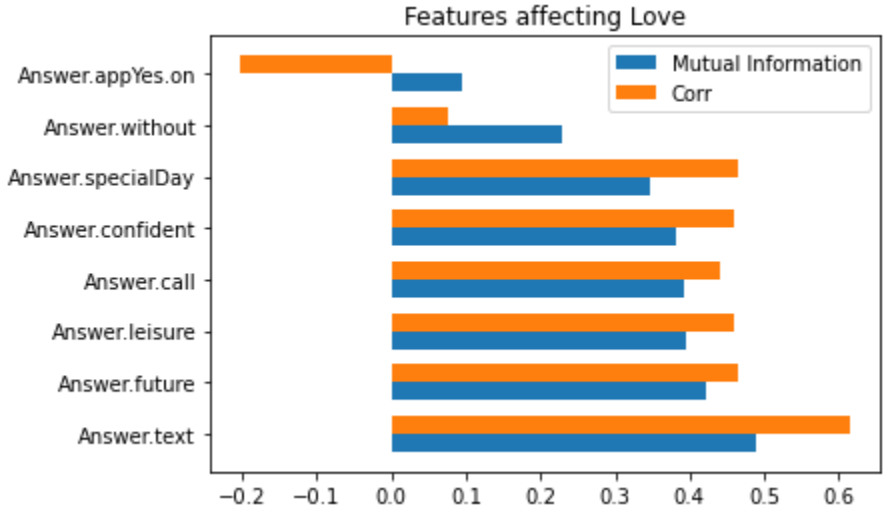
Detail results:

Features	Mutual	Correlations	Chi-square

	Information		p-value
Answer.specialDay	0.493240	0.617947	1.686107e-18
Answer.leisure	0.472726	0.512281	4.018918e-18
Answer.future	0.401133	0.443235	1.665181e-17
Answer.love	0.381117	0.460289	4.355942e-16
Answer.call	0.335060	0.272481	8.715662e-08
Answer.text	0.317113	0.273128	7.406160e-06
Answer.without	0.297877	0.005659	2.504705e-02
Answer.appYes.on	0.071308	-0.233246	5.690819e-03

(Full results can be found in the code)
From the above result, we can see that all activities including Celebrating Special Days, Leisure, Talking about Future, Feeling love, Call, Text have positive effect on the Confident about the relationship, or the more frequently these activities are, the more confident users are in their relationship. The time spent alone in the relationship also shows its significance, but Correlations show none, which suggests this feature relates to Confident in non-linear terms. We also found that Not using the App seems to improve the Confident about the relationship, but correlation is 2-way, so Not confident in their relationship also requires/boosts their usage on the app.

Graphs for features affecting Feelings about current Relationship:



Detail Results

Features	Mutual Information	Correlations	Chi-square p-value
Answer.specialDay	0.348109	0.465530	1.176634e-07
Answer.leisure	0.395366	0.459270	1.134061e-08
Answer.future	0.422186	0.465016	3.283094e-18
Answer.confident	0.381117	0.460289	4.355942e-16
Answer.call	0.393729	0.441877	5.923039e-12
Answer.text	0.489674	0.614418	3.442441e-26
Answer.without	0.228019	0.075869	7.330261e-01
Answer.appYes.on	0.094726	-0.201559	6.617396e-04

(Full results can be found in the code)

From the above result, we can see that all activities including Leisure, Talking about Future, Feeling confident, Call, Text have positive effect on the Feeling about the relationship, or the more frequently these activities are, the more love users feel in their relationship. The time spent alone in the relationship also shows its significance, but Correlations show none, which suggests this feature relates to Love in non-linear terms. We also found that Not using the App seems to improve the Confident about the relationship, but as mentioned above, it can be 2-way.

b. Linear Regression Model

We first choose appropriate features to put in our model. We need to handpick the features, based on the correlations above, as well as their correlations with each other, to avoid multicolinearity. The correlation tables of mentioned variables are as following:

	Answer.specialDay	Answer.leisure	Answer.future	Answer.love	Answer.call	Answer.text	Answer.without	Answer.appYes.on
Answer.specialDay	1.000000	0.629621	0.544899	0.465530	0.363188	0.360469	-0.004970	-0.027509
Answer.leisure	0.629621	1.000000	0.393422	0.459270	0.383983	0.389097	0.084442	-0.061389
Answer.future	0.544899	0.393422	1.000000	0.465016	0.447070	0.467446	0.103540	-0.000174
Answer.love	0.465530	0.459270	0.465016	1.000000	0.441877	0.614418	0.075869	-0.201559
Answer.call	0.363188	0.383983	0.447070	0.441877	1.000000	0.429291	0.047215	0.026151
Answer.text	0.360469	0.389097	0.467446	0.614418	0.429291	1.000000	0.124226	-0.001100
Answer.without	-0.004970	0.084442	0.103540	0.075869	0.047215	0.124226	1.000000	0.096686
Answer.appYes.on	-0.027509	-0.061389	-0.000174	-0.201559	0.026151	-0.001100	0.096686	1.000000

The features we choose are: Answer.specialDay, Answer.call, Answer.without, and Answer.appYes.on. The Linear Regression model is then fitted, with these features. The result on the whole data is R-squared equal 0.43, with mean-squared-error equal to 3.39.
The coefficients are:

- Answer.SpecialDay: 0.56971961
- Answer.call: 0.05361383
- Answer.without: 0.02632814
- Answer.appYes.on: -1.09142708

Based on the above coefficients, we can see that even in a linear regression model, Celebrating SpecialDay adds the most values to Confident, compared to other features. Answer.appYes.on, to our surprise, contributes heavily to the overall model, with the change from No to Yes can reduce the score scale by 1.

Similarly, we base on the below correlation table to detect the features for the model

	Answer.text	Answer.future	Answer.leisure	Answer.call	Answer.specialDay	Answer.without	Answer.appYes.on
Answer.text	1.000000	0.467446	0.389097	0.429291	0.360469	0.124226	-0.001100
Answer.future	0.467446	1.000000	0.393422	0.447070	0.544899	0.103540	-0.000174
Answer.leisure	0.389097	0.393422	1.000000	0.383983	0.629621	0.084442	-0.061389
Answer.call	0.429291	0.447070	0.383983	1.000000	0.363188	0.047215	0.026151
Answer.specialDay	0.360469	0.544899	0.629621	0.363188	1.000000	-0.004970	-0.027509
Answer.without	0.124226	0.103540	0.084442	0.047215	-0.004970	1.000000	0.096686
Answer.appYes.on	-0.001100	-0.000174	-0.061389	0.026151	-0.027509	0.096686	1.000000

The features we chose are: Answer.text, Answer.without, Answer.appYes.on. The result on Linear Regression gives out the mean-squared-error of 4.02 with R2 equal to 0.42. The coefficients obtained are:

- Answer.text: 0.60090057
- Answer.without: 0.02040345
- Answer.appYes.on: -1.07469723

The Answer.text has strong influence, with one point in this feature increasing 0.6 in Passion. Answer.appYes.on also gives heavy reduction on Love, with deduction of 1 when users use Apps supporting their Relationship. Once again, the Answer.without gave very insignificant linear change in Passion, which we will explore more in the future.

VI. Conclusion

In this project, we explore the behaviors of users relating to their relationship. We check which factors strengthen or weaken their certainty as well as their feelings toward their love. We obtain the data from the Mturk-platform survey to obtain data, which, to some

extent, can represent public data. We explore that the factors relating to daily behaviors, like Texting, Calling, Celebrating Special Days, etc. can help strengthen the feeling and confidence in a relationship. We also found that Using too many Love Apps can reduce these feelings in a relationship. Another feature is Spending time alone without a partner contributes in non-linear terms to the factors we explore, which can be subject to the future. Finally, some regressions are fitted into the data, to quantify how the feelings will change as one of the factors shift. Although the result is only decent, we have a better look at some behaviors' significance toward feeling and confidence in a relationship.

VII. Reference

[1] Alvin Powell, 2018, *When love and science double date*, Health and Medicine, The Harvard Gazelle

[2] Michelle Duda, 2014, *What Kills Love? Factors Influencing the Ending of Love in Romantic Relationships*, Illinois State University

[3] Inga Großmann ,André Hottung,Artus Krohn-Grimberghe, 2019, *Machine learning meets partner matching: Predicting the future relationship quality based on personality traits* ,
https://doi.org/10.1371/journal.pone.0213569

[5] Beverley Fehr, 1988, *Prototype Analysis of the Concepts of Love and Commitment*

[6] Samantha Joel, Paul W. Eastwick, Colleen J. Allison, View ORCID ProfileXimena B. Arriaga, Zachary G. Bake. 2020. *Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies*

[7] Agishtein, P., & Brumbaugh, C. (2013). *Cultural variation in adult attachment: The impact of ethnicity, collectivism, and country of origin*. Journal of Social, Evolutionary, and Cultural Psychology, 7(4), 384-405. http://dx.doi.org/10.1037/h0099181

[8] Beste, S. A., Bergner, R. M., & Nauta, M. M. (2003). *What keeps love alive? An empirical investigation*. Family Therapy, 30(3), 125-141. Billings, A. (1979). Conflict resolution in distressed and nondistressed married couples. Journal of Consulting and Clinical Psychology, 47(2), 368-376. doi: 10.1037/0022- 006X.47.2.368

[9] Pearson, Karl (1900). "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling" . Philosophical Magazine. Series 5. 50 (302): 157–175. doi:10.1080/14786440009463897.

[10] Lund, B. (2019) *Transforming Ordinal Predictors to Numeric for Binary Logistic Models*, Corpus ID: 202635770

[11] Croxton, Frederick Emory; Cowden, Dudley Johnstone; Klein, Sidney (1968) Applied General Statistics, Pitman. ISBN 9780273403159 (page 625)

[12] David A. Freedman (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 26. A simple regression equation has on the right hand side an intercept and an explanatory variable with a slope coefficient. A multiple regression e right hand side, each with its own slope coefficient