Machine Learning for Relationship Outcome Prediction

Christopher Labiak, Yao Liu, Matthew Kliemann, Kaustubh Srivastava, Yao Xiao

University of Michigan, College of Engineering, Computer Science Engineering

UNIVERSITY OF MICHIGAN

Abstract

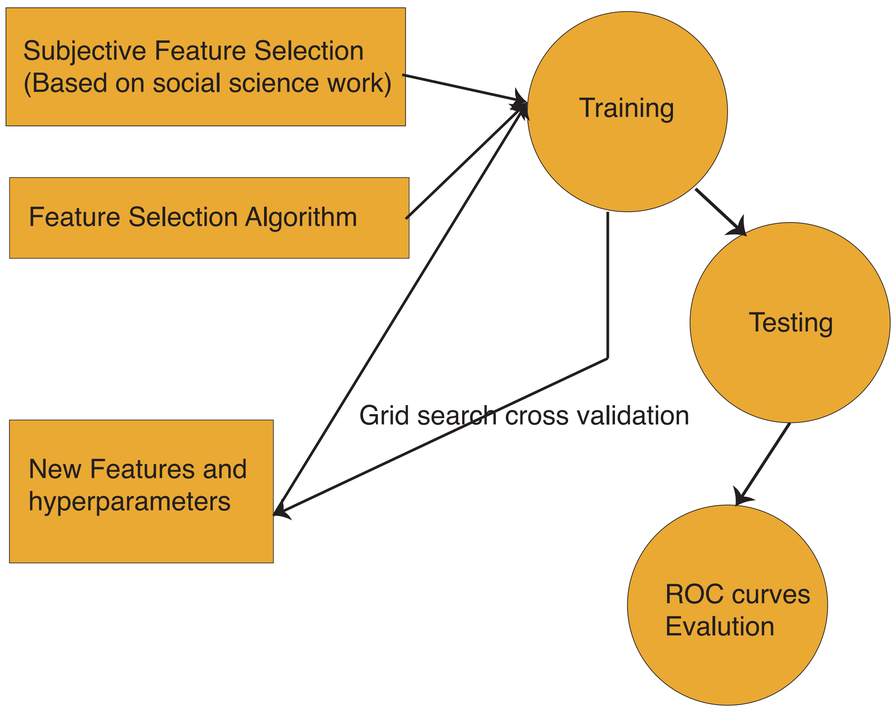
- Using the How Couples Meet and Stay Together survey from Stanford
- Predict the 4 year outcome of personal relationships that were active at beginning of survey
- Understand most significant contributing factors to relationship success or failure
- Identifying incompatible relationships at inception

Data Preprocessing

- Manually label
  - Positive
    - broke up in the middle (optimistic prediction)
  - Negative
    - stayed together for 4-year period
- Different treatment of continuous and categorical data
  - Categorical data
    - One hot encoded to account for lack of relationship
    - Missing values: most frequent
    - Keep refusal answers as additional category
  - Continuous data
    - Rescaled between 0 and 1
    - Missing values filled using median
- Data size
  - 1873 examples * 143 features after preprocessing

Method Pipelines

- 80-20 testing-training split
- 5-fold validation
- Multiple training models: SVM, K-nearest neighbors, Naive Bayes, Regression Classifier, Neural Network (backpropagation)

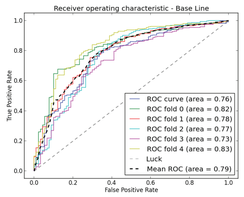


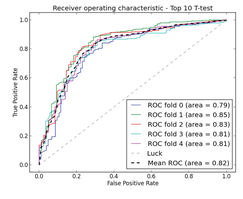
Subjective Features

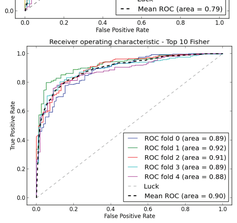
- Equality of earnings reduces the likelihood of breakup for same-sex couples, while it increases the likelihood of
- Couples in which the respondent has higher years of education are less likely to experience a breakup
- Being married or in a domestic partnership significantly decreases the likelihood of breakup
- The likelihood of breakup decreases with higher household incomes
- Couples who met through friends had higher than average breakup rates
- Couples who met in primary or secondary school and met in church have substantially lower couple dissolution rates
- The internet rises as increasing social intermediary to find partners

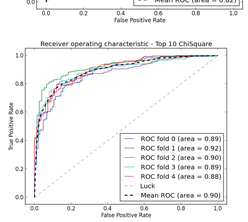
Experiments

- Feature Selection with cross validated gaussian-SVM









- Weight Vector of Toy Logistic Regression Classifier

Feature	Weight (Truth)
Couple is cohabitating	1.79
Attend same high school	-0.44
Not married	-1.20
High Parental Approval	1.03
Divorced	-1.21
Rented living quarters for cash	-0.42
Age when romantic relationship began	-0.04
Met using internet service	0.22
Met using internet service or online or offline	-0.52
Met online of offline	-0.21

- Chi-Squared Confusion Matrix with gaussian SVM

	Success Prediction	Fail Prediction
Actual Success	80	44
Actual Failure	58	380

Test Accuracy 81.85%

Conclusions

- Similar prediction accuracy of approximately 80% among all models
- Challenges
  - Limited dataset size
  - Limited subjective and emotional data about partners and relationship beyond demographic information
  - Insufficient amount of data to explore difference between marital relationship and “dating relationship”
- Domain knowledge overlaps with feature selection algorithm
- Predict your relationship outcome tool on web may be useful

References

- Gottman, John M., and Robert W. Levenson. "The Timing of Divorce: Predicting When a Couple Will Divorce Over a 14-Year Period." Journal of Marriage and Family 62.3 (2000): 737-45. Web.
- Pedregosa, and Et Al. "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research 12 (2011): 2825-830. Web.
- Rosenfeld, Michael J. How Couples Meet and Stay Together. Stanford University, 2014. Web.