Chapter 6

Cross-Disciplinary ML Research is like Happy Marriages: Five Strengths and Two Examples

Meng Jiang
University of Notre Dame

Top Strengths in ML+X Collaboration

Cross-disciplinary research refers to research and creative practices that involve two or more academic disciplines (Jeffrey 2003; Karniouchina, Victorino, and Verma 2006). These activities may range from those that simply place disciplinary insights side by side to much more integrative or transformative approaches (Aagaard-Hansen 2007; Muratovski 2011). Cross-disciplinary research matters, because (1) it provides an understanding of complex problems that require a multifaceted approach to solve; (2) it combines disciplinary breadth with the ability to collaborate and synthesize varying expertise; (3) it enables researchers to reach a wider audience and communicate diverse viewpoints; (4) it encourages researchers to confront questions that traditional disciplines do not ask while opening up new areas of research; and (5) it promotes disciplinary self-awareness about methods and creative practices (Urquhart et al. 2011; O'Rourke, Crowley, and Gonnerman 2016; Miller and Leffert 2018).

One of the most popular cross-disciplinary research topics/programs is *Machine Learning + X* (or *Data Science + X*). Machine learning (ML) is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns, and make decisions with minimal human intervention. ML has been used in a variety of applications (Murthy 1998), such as email filtering and computer vision; however, most applications still fall in the domain of computer science and engineering. Recently, the power of *ML+X*, where *X* can be *any* other discipline (such as physics, chemistry,

biology, sociology, and psychology), is well recognized. ML tools can reveal profound insights hiding in ballooning datasets (Kohavi et al. 1994; Pedregosa et al. 2011; Kotsiantis 2012; Mullainathan and Spiess 2017).

However, cross-disciplinary research, which *ML+X* is part of, is challenging. Collaborating with investigators outside one's own field requires more than just adding a co-author to a paper or proposal. True collaborations will not always be without conflict—lack of information leads to misunderstandings. For example, ML experts would have little domain knowledge in the field of X; and researchers in X might not understand ML either. The knowledge gap limits the progress of collaborative research.

So how can we start and manage successful cross-disciplinary research? What can we do to facilitate collaborative behaviors? In this essay, I will compare cross-disciplinary ML research to "happy marriages," discussing some characteristics they share. Specifically, I will present the top strengths of conducting cross-disciplinary ML research and give two examples based on my experience of collaborating with historians and psychologists.

Marriage is one of the most common "collaborative" behaviors. Couples expect to have happy marriages, just like collaborators expect to have successful project outcomes (Robinson and Blanton 1993; Pettigrew 2000; Xu et al. 2007). Extensive studies have revealed the top strengths of happy marriages (DeFrain and Asay 2007; Gordon and Baucom 2009; Prepare/Enrich, n.d.), which can be reflected in cross-disciplinary ML research. Here I focus on five of them:

1. Collaborators ("partners" in the language of marriage) are satisfied with *communication*.

2. Collaborators *feel very close* to each other.

3. Collaborators *discuss their problems* well.

4. Collaborators *handle their differences* creatively.

5. There is a *good balance of time alone* (i.e., individual research work) *and together* (meetings, discussions, etc).

First of all, communication is the exchange of information to achieve a better understanding; and collaboration is defined as the process of working together with another person to achieve an end goal. Effective collaboration is about sharing information, knowledge, and resources to work together through satisfactory communication. Ineffectiveness or lack of communication is one of the biggest challenges in *ML+X* collaboration.

Second, researchers in different disciplines meet different challenges through the process of collaboration. Making the challenges clear to understand and finding solutions together is the core of effective collaboration.

Third, researchers in different disciplines can collaborate only when they recognize mutual interest and feel that the research topics they have studied in depth are very close to each other. Collaborators must be interested in solving the same, big problem.

Fourth, collaborators must embrace their differences on concepts and methods and take advantage of them. For example, one researcher can introduce a complementary method to the mix of other methods that the collaborator has been using for a long time; or one can have a new, impactful dataset and evaluation method to test the techniques proposed by the other.

Fifth, in strong collaboration, there is a balance between separateness and togetherness. Meetings are an excellent use of time for having integrated perspectives and productive discourse around

difficult decisions. However, excessive collaboration happens when researchers are depleted by too many meetings and emails. It can lead to inefficient, unproductive meetings. So it is important to find a balance.

Next, I, as a computer scientist and ML expert, will discuss two *ML+X* collaborative projects. ML experts bring mathematical modeling and computational methods for mining knowledge from data. The solutions usually have good generalizability; however, they still need to be tailored for specialized domains or disciplines.

Example 1: ML + History

The history professor Liang Cai and I have collaborated on an international research project titled "Digital Empires: Structured Biographical and Social Network Analysis of Early Chinese Empires." Dr. Cai is well known for her contributions to the fields of early Chinese Empires, Classical Chinese thought (in particular, Confucianism and Daoism), digital humanities, and the material culture and archaeological texts of early China (Cai 2014). Our collaboration explores how digital humanities expand the horizon of historical research and help visualize the research landscape of Chinese history. Historical research is often constrained by sources and the human cognitive capacity for processing them. ML techniques may enhance historians' abilities to organize and access sources as they like. ML techniques can even create new kinds of sources at scale for historians to interpret.

> "The historians pose the research questions and visualize the project," said Cai. "The computer scientists can help provide new tools to process primary sources and expand the research horizon."

We conducted a structured biographical analysis to leverage the development of machine learning techniques, such as neural sequence labeling and textual pattern mining, which allowed classical sources of Chinese empires to be represented in an encoded way. The project aims to build a digital biographical database that sorts out different attributes of all recorded historical actors in available sources. Breaking with traditional formats, ML+History creates new opportunities and augments our way of understanding history.

First, it helps scholars, especially historians, change their research paradigm, allowing them to generalize their arguments with sufficient examples. ML techniques can find all examples in the data where manual investigation may miss some. Also, abnormal cases can indicate a new discovery. As far as early Chinese empires are concerned, ML promises to automate mining and encoding all available biographical data, which allows scholars to change the perspective from one person to a group of persons with shared characteristics, and to shift from analyzing examples to relating a comprehensive history. Therefore, scholars can identify general trends efficiently and present an information-rich picture of historical reality using ML techniques.

Second, the structured data produced by ML techniques revolutionize the questions researchers ask, thereby changing the research landscape. Because of the lack of efficient tools, there are numerous interesting questions scholars would like to ask but cannot. For example, the geographical mobility of historical actors is an intriguing question for early China, the answer to which would show how diversified regions were integrated into a unified empire. Nevertheless, an individual historian cannot efficiently process the massive amount of information preserved in the sources. With ML techniques, we can generate fact tuples to sort out original geographical places of all available historical actors and provide comprehensive data for historians to analyze.
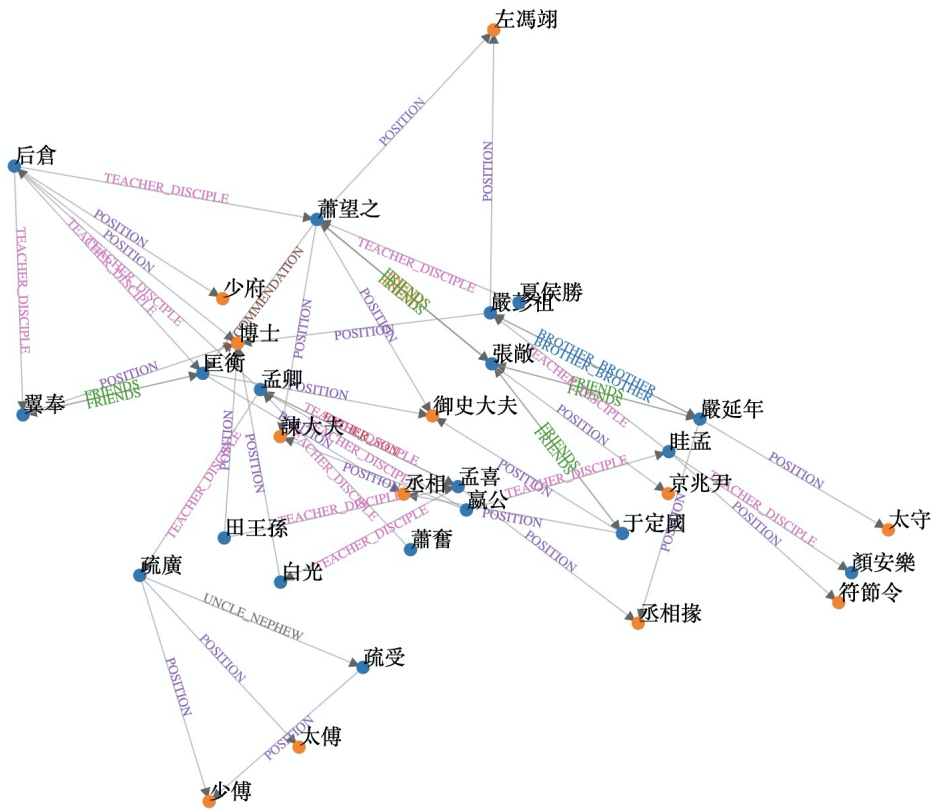
Figure 6.1: The graph presents a visual of the social network of officials who served in the government about 2,000 years ago in China. The network describes their relationships and personal attributes.

Patterns Mined by ML Tech	Extracted Relations
$PER_X ...從 $PER_Y 受 $KLG *$PER_X was taught by $PER_Y on $KLG (knowledge)*	(張禹, 施讎, 易) (施讎, 田王孫, 易) (眭弘, 嬴公, 春秋)
$PER_X ... 事 $PER_Y *$PER_X was taught/mentored by $PER_Y*	(司馬相如, 孝景帝) (尹齊, 張湯)
$PER_X ... 授 $PER_Y *$PER_X taught $PER_Y*	(孟喜, 后蒼、疏廣) (王式, 龔舍)
$PER ... $LOC 人也 *$PER place_of_birth $LOC*	(張敞, 河東平陽) (彭越, 昌邑)
$PER 遷 $TIT *$PER job_title $TIT*	(朱邑, 北海太守) (甘延壽, 遼東太守)
$PER 至 $TIT *$PER job_title $TIT*	(歐陽生, 御史大夫) (孟卿, 中山中尉)
$PER 為 $TIT *$PER job_title $TIT*	(伏生, 秦博士) (司馬相如, 武騎常侍)

Table 6.1: Examples of Chinese Text Extraction Patterns

Third, the project revolutionizes our reading habits. Large datasets mined from primary sources will allow scholars to combine long-distant reading with original texts. The macro picture generated from data will aid in-depth analysis of the event against its immediate context. Furthermore, graphics of social networks and common attributes of historical figures will change our reading habits, transforming linear storytelling to accommodate multiple narratives (see the above figure).

Researchers from the two sides develop collaboration through the project step by step, just like developing a relationship for marriage. Ours started at a faculty gathering from some random chat about our research. As the historian is open-minded to ML technologies and the ML expert is willing to create broader impact, we brainstormed ideas that would not have developed without taking care of the five important points:

1. *Communication:* With our research groups, we started to meet frequently at the beginning. We set up clear goals at the early stage, including expected outcomes, publication venues, and joint proposals for funding agencies, such as the National Endowment for the Humanities (NEH) and Notre Dame seed grant funding. Our research groups met almost twice a week for as long as three weeks.

2. *Feel very close to each other:* Besides holding meetings, we exchanged our instant messenger accounts so we could communicate faster than email. We created Google Drive space to share readings, documents, and presentation slides. We found many tools to create "tight relationships" between the groups at the beginning.

3. *Discuss their problems well:* Whenever we had misunderstandings, we discussed our prob-

lems. Historians learned about what a machine does, what a machine can do, and generally how a machine works toward the task. ML people learned what is interesting to historians and what kind of information is valuable. We hold the principle that as the problems exist, they make sense; any problem any other encounters is worth a discussion. We needed to solve problems together from the moment they became our problems.

4. *Handle their differences creatively:* Historians are among the few who can read and write in classical Chinese. Classical Chinese was used as the written language from over 3,000 years ago to the early 20th century. Since then, mainland China has used either Mandarin (simplified Chinese) or Cantonese, while Taiwan has used traditional Chinese. None is similar to classical Chinese at all. In other words, historians work on a language that no ML experts here, even those who speak modern Chinese, can understand. So we handle our language differences "creatively" by using the translated version as the intermediate medium. Historians have translated history books in classical Chinese into simplified Chinese so we can read the simplified version. Here, the idea is to let the machine learning algorithms read both versions. We find that information extraction (i.e., finding relations from text) and machine translation (i.e., from classical Chinese to modern Chinese) can mutually enhance each other, which turns out to be one of our novel technical contributions to the field of natural language processing.

5. *Good balance of time alone and together:* After the first month, since the project goal, datasets, background knowledge, and many other aspects were clear in both sides' minds, we had regular meetings in a less intensive manner. We met twice or three times a month so that computer science students could focus on developing machine learning algorithms, and only when significant progress was made or expert evaluation was needed would we schedule a quick appointment with Prof. Liang Cai.

So far, we have published peer-reviewed papers on the topic of information extraction and entity retrieval in classical Chinese history books using ML (Ma et al. 2019; Zeng et al. 2019). We have also submitted joint proposals with the above work as preliminary results to NEH.

Example 2: ML + Psychology

I am working with Drs. Ross Jacobucci and Brooke Ammerman in psychology to apply ML to understand mental health problems and suicidal intentions. Suicide is a serious public health problem; however, suicides are preventable with timely, evidence-based interventions. Social media platforms have been serving users who are experiencing real-time suicidal crises with hopes of receiving peer support. To better understand the helpfulness of peer support occurring online, we characterize the content of both a user's post and corresponding peer comments occurring on a social media platform and present an empirical example for comparison. We have designed a new topic-model-based approach to finding topics of users and peer posts from the social media forum data. The key advantages include: (i) modeling both the generative process of each type of corpora (i.e., user posts and peer comments) and the associations between them, and (ii) using phrases, which are more informative and less ambiguous than words alone, to represent social media posts and topics. We evaluated the method using data from Reddit's r/SuicideWatch community.
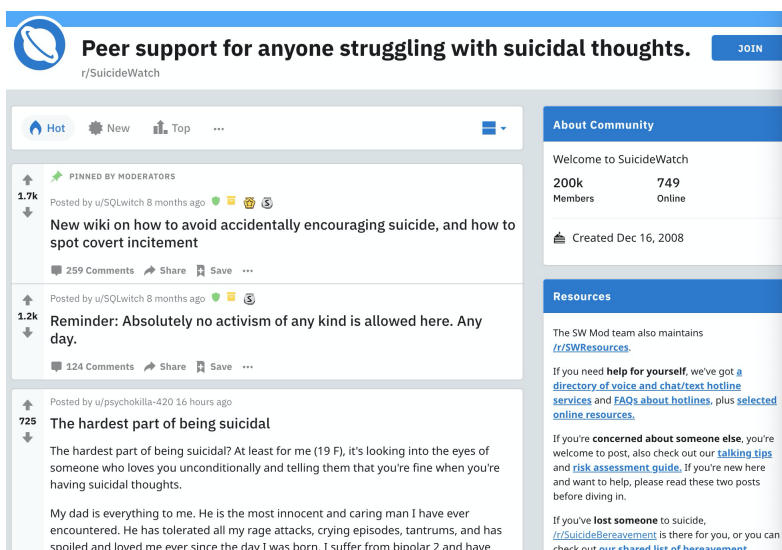
Figure 6.2: Screenshot of r/SuicideWatch on Reddit.

We examined how the topics of user and peer posts were associated and how this information influenced the perceived helpfulness of peer support. Then, we applied structural topic modeling to data collected from individuals with a history of suicidal crisis as a means to validate findings. Our observations suggest that effective modeling of the association between the two lines of topics can uncover helpful peer responses to online suicidal crises, notably providing the suggestion of pursuing professional help. Our technology can be applied to "paired" corpora in many applications such as tech support forums and question-answering sites.

This project started from a talk I gave at the psychology graduate seminar. The fun thing is that Dr. Jacobucci was not able to attend the talk. Another psychology professor who attended my talk asked constructive questions and mentioned my research to Dr. Jacobucci when they met later. So Dr. Jacobucci dropped me an email, and we had coffee together. Cross-disciplinary research often starts from something that sounds like developing a relationship. Because, again, the psychologists are open-minded to ML technologies and the ML expert is willing to create broader impact, we successfully brainstormed ideas when we had coffee, but this would not have developed into long-term collaboration without the following efforts: (1) Communicate intensively between research groups at the early stage. We had multiple meetings a week to make the goals clear. (2) Get students involved in the process. When my graduate student received more and more advice from the psychology professors and students, the connections between the two groups became stronger. (3) Discuss the challenges in our fields very well. We analyzed together whether machine learning would be capable of addressing the challenges in mental health. We also analyzed whether domain experts could be involved in the loop of machine learning algorithms. (4) Handle our differences. We separately presented our research and then found times to work together to put sets of slides together based on one common vision and goal. (5) After the first month, only hold meetings when discussion is needed or there is an approaching deadline

for either paper or proposal.

We have enjoyed our collaboration and the power of cross-disciplinary research. Our joint work is under review at *Nature Palgrave Communications*. We have also submitted joint proposals to NIH with this work as preliminary results (Jiang et al. 2020).

Conclusions

In this essay, I used a metaphor comparing cross-disciplinary ML research to "happy marriages." I discussed five characteristics they share. Specifically, I presented the top strengths of producing successful cross-disciplinary ML research: (1) Partners are satisfied with *communication*. (2) Partners *feel very close* to each other. (3) Partners *discuss their problems* well. (4) Partners *handle their differences* creatively. (5) There is a *good balance* of time alone (i.e., individual research work) *and together* (meetings, discussions, etc). While every project is different and will produce its own challenges, my experience of collaborating with historians and psychologists according to the happy marriage metaphor suggests that it is a simple and strong paradigm that could help other interdisciplinary projects develop into successful, long-term collaborations.

References

Aagaard‐Hansen, Jens. 2007. "The Challenges of Cross‐Disciplinary Research." *Social Epistemology* 21, no. 4 (October-December): 425–38. `https://doi.org/10.1080/02691720701746540`.

Cai, Liang. 2014. *Witchcraft and the Rise of the First Confucian Empire*. Albany: SUNY Press.

DeFrain, John, and Sylvia M. Asay. 2007. "Strong Families Around the World: An Introduction to the Family Strengths Perspective." *Marriage & Family Review* 41, no. 1–2 (August): 1–10. `https://doi.org/10.1300/J002v41n01_01`.

Gordon, Cameron L., and Donald H. Baucom. 2009. "Examining the Individual Within Marriage: Personal Strengths and Relationship Satisfaction." *Personal Relationships* 16, no. 3 (September): 421–435. `https://doi.org/10.1111/j.1475-6811.2009.01231.x`.

Jeffrey, Paul. 2003. "Smoothing the Waters: Observations on the Process of Cross-Disciplinary Research Collaboration." *Social Studies of Science* 33, no. 4 (August): 539–62.

Jiang, Meng, Brooke A. Ammerman, Qingkai Zeng, Ross Jacobucci, and Alex Brodersen. 2020. "Phrase-Level Pairwise Topic Modeling to Uncover Helpful Peer Responses to Online Suicidal Crises." *Humanities and Social Sciences Communications* 7: 1–13.

Karniouchina, Ekaterina V., Liana Victorino, and Rohit Verma. 2006. "Product and Service Innovation: Ideas for Future Cross-Disciplinary Research." *The Journal of Product Innovation Management* 23, no. 3 (May): 274–80.

Kohavi, Ron, George John, Richard Long, David Manley, and Karl Pfleger. 1994. "MLC++: A Machine Learning Library in C++." In *Proceedings of the Sixth International Conference on Tools with Artificial Intelligence*, 740–3. N.p.: IEEE. `https://doi.org/10.1109/TAI.1994.346412`.

Kotsiantis, S.B. 2012. "Use of Machine Learning Techniques for Educational Proposes [*sic*]: a Decision Support System for Forecasting Students' Grades." *Artificial Intelligence Review* 37, no. 4 (May): 331–44. `https://doi.org/10.1007/s10462-011-9234-x`.

Ma, Yihong, Qingkai Zeng, Tianwen Jiang, Liang Cai, and Meng Jiang. 2019. "A Study of Person Entity Extraction and Profiling from Classical Chinese Historiography." In *Proceedings of the 2nd International Workshop on EntitY REtrieval*, edited by Gong Cheng, Kalpa Gunaratna, and Jun Wang, 8–15. N.p.: International Workshop on EntitY REtrieval. `http://ceur-ws.org/Vol-2446/`.

Miller, Eliza C. and Lisa Leffert. 2018. "Building Cross-Disciplinary Research Collaborations." *Stroke* 49, no. 3 (March): e43-e45. `https://doi.org/10.1161/strokeaha.117.020437`.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine learning: an applied econometric approach." *Journal of Economic Perspectives* 31, no. 2 (spring): 87–106. `https://doi.org/10.1257/jep.31.2.87`.

Muratovski, Gjoko. 2011. "Challenges and Opportunities of Cross-Disciplinary Design Education and Research." In *Proceedings from the Australian Council of University Art and Design Schools (ACUADS) Conference: Creativity: Brain—Mind—Body*, edited by Gordon Bull. Canberra, Australia: ACAUDS Conference. `https://acuads.com.au/conference/article/challenges-and-opportunities-of-cross-disciplinary-design-education-and-research/`.

Murthy, Sreerama K. 1998. "Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey." *Data Mining and Knowledge Discovery* 2, no. 4 (December): 345–89. `https://doi.org/10.1023/A:1009744630224`.

O'Rourke, Michael, Stephen Crowley, and Chad Gonnerman. 2016. "On the Nature of Cross-Disciplinary Integration: A Philosophical Framework." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 56 (April): 62–70. `https://doi.org/10.1016/j.shpsc.2015.10.003`.

Pedregosa, Fabian et al. 2011. "Scikit-learn: Machine Learning in Python." *The Journal of Machine Learning Research* 12: 2825–30. `http://www.jmlr.org/papers/v12/pedregosa11a.html`.

Pettigrew, Simone F. 2000. "Ethnography and Grounded Theory: a Happy Marriage?" In *Association for Consumer Research Conference Proceedings,* edited by Stephen J. Hoch and Robert J. Meyer, 256–60. Provo, UT: Association for Consumer Research. `https://www.acrwebsite.org/volumes/8400/volumes/v27/`.

Prepare/Enrich. N.d. "National Survey of Marital Strengths." Prepare/Enrich (website). Accessed January 17, 2020. `https://www.prepare-enrich.com/pe_main_site_content/pdf/research/national_survey.pdf`.

Robinson, Linda C. and Priscilla W. Blanton. 1993. "Marital Strengths in Enduring Marriages." *Family Relations: An Interdisciplinary Journal of Applied Family Studies* 42, no. 1 (January): 38–45. `https://doi.org/10.2307/584919`.

Urquhart, R., E. Grunfeld, L. Jackson, J. Sargeant, and G. A. Porter. 2013. "Cross-Disciplinary Research in Cancer: an Opportunity to Narrow the Knowledge–Practice Gap." *Current Oncology* 20, no. 6 (December): e512–e521. `https://doi.org/10.3747/co.20.1487`.

Xu, Anqi, Xiaolin Xie, Wenli Liu, Yan Xia, and Dalin Liu. 2007. "Chinese Family Strengths and Resiliency." *Marriage & Family Review* 41, no. 1–2 (August): 143–64. `https://doi.org/10.1300/J002v41n01_08`.

Zeng, Qingkai, Mengxia Yu, Wenhao Yu, Jinjun Xiong, Yiyu Shi, and Meng Jiang. 2019. "Faceted Hierarchy: A New Graph Type to Organize Scientific Concepts and a Construction Method."

In *Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13)*, edited by Dmitry Ustalov, Swapna Somasundaran, Peter Jansen, Goran Glavaš, Martin Riedl, Mihai Surdeanu, and Michalis Vazirgiannis, 140–50. Hong Kong: Association for Computational Linguistics. `https://doi.org/10.18653/v1/D19-5317`.