USCDornsife

Psych 626: Topics in Computational Social Sciences

Fall—TH—2-6
Location: BCI 266

Instructor: Morteza Dehghani
Office: SGM 607
Office Hours: M 10-12, or by appt.
Contact Info: mdehghan@usc.edu

Course Description

Topics in computational social sciences focuses on applications of big-data methods (machine learning techniques, natural language processing and network analysis), guided by psychological theories, for identifying various social and cognitive properties evident in human related big data. In this course, we will survey state-of-the-art techniques, and applications of such techniques, for investigating various aspects of human cognition. The intended audience for this course is psychology and computer science PhD students, and more broadly graduate students in social sciences, who are interested in using machine-learning techniques for analysis of data. Also, this course may be of interest to PhD students in communications and the business school.

Learning Objectives

This course is designed to survey current state of research in computational social sciences. In order to achieve this objective, each week several papers/books will be read and presented by students. Also, there will be a final project and written report.

Prerequisite(s): Instructor permission
Recommended Preparation: Psych 625 or a similar course

Course Notes

Students are not allowed to use laptops or smartphones during class, unless used for class presentations. Homework assignments will be posted on Blackboard. Students are also highly encouraged to use the course forum on Blackboard.

Technological Proficiency and Hardware/Software Required

The project for the class involves programming. Students are recommended to have taken a programming course before enrolling in this course.

Description and Assessment of Assignments

a. Paper presentation. Each student will present a set of papers related to one of the topics discussed in class.
b. Reaction paragraphs. Students are asked to write a short note, one or two paragraphs in length, about their reaction to the reading assignments of the week. These can be a quick summary of the material, comments about the subject area, or a critique of a particular theory or experiment. I will read these paragraphs before each class, and will use them to guide the discussion in class.
c. Class Project. This class is project oriented. The goal of the project is for students to get experience in applying big-data methods for analyzing behavioral data. This will include a project proposal presentation, three project update presentations, final project presentation, and a report. For project proposals, students will present a problem and a data collection method and/or dataset for which they want to analyze using the methods discussed in class. Each presentation should be about 10-15mins. The goal of the project update presentations

is to inform the class about the state of the project and brainstorm with other students on how to solve the remaining issues. Each update presentation should be around 10 minutes. For the final project presentation, each student/group will give a 15-20min presentation on their project. Students are expected to spend at least 80 hours working on their final project. The project report will be 10-15 pages.

d. Review Paragraphs: Students will provide reviews for other students' projects. The reviews for the project proposal and project updates can each be about a paragraph in length and should address the strength and weaknesses of the discussed projects and other potential sources of related work. Each student is asked to read the final report of two other groups and provide comprehensive reviews of the final projects as well. These reviews will be sent to the authors, and the authors will have a chance to respond to the reviews.

Grading Breakdown

Assignment	% of Grade
Project Proposal	10
Project Update 1	5
Project Update 2	5
Project Update 3	5
Project Presentation	10
Final Project Report	15
Reviews	10
Participation	15
Paper Presentations	15
Reaction Paragraphs	10
TOTAL	100

Assignment Submission Policy

The reaction paragraphs need to be submitted through Blackboard before 9:00am of the day the corresponding readings are due. The final project report and the reviews need to be submitted through Blackboard on the day of the final.

Course Schedule:

The following schedule is tentative and may change during the semester (* indicates optional readings).

- Week 1: Introduction to Computational Social Sciences
  - Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science*, *346*(6213), 1063-1064.
  - Lazer D, et al. (2009) Computational social science. *Science,* 323(5915):721–723.
  - Yarkoni, T. (2012). Psychoinformatics New Horizons at the Interface of the Psychological and Computing Sciences. *Current Directions in Psychological Science*, 21(6), 391-397.
  - King, G. (2011). Ensuring the data-rich future of the social sciences. *Science*, 331(6018), 719:721.

- Week 2: Discussions related to Data
  - Borgman, C. L. (2015). *Big data, little data, no data: scholarship in the networked world*. MIT Press. Required chapters: 1, 2, 5, 6, 7 & 8
  - Wienberg, C. and Gordon, A. (2015) Insights on Privacy and Ethics from the Web's Most Prolific Storytellers. *Proceedings of ACM Web Science* 2015, June 28 - July 1, 2015, Oxford,

UK.
    - Watch in class: Friends You Haven't Met Yet
    - *Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA*, *309*(13), 1351-1352

- Week 3: Text Analysis in Social Sciences – Word Count 1
    - Pennebaker, J. W. (2011). *The secret life of pronouns: How our words reflect who we are.* New York, NY: Bloomsbury Press.

- Week 4: Text Analysis in Social Sciences – Word Count 2, Project Proposals
    - Back, M. D., Küfner, A. C., & Egloff, B. (2010). The emotional timeline of September 11, 2001. *Psychological Science*, 21(10), 1417:1419.
    - Pury, C. L. (2011). Automation can lead to confounds in text analysis Back, Küfner, and Egloff (2010) and the Not-So-Angry Americans. *Psychological science*.
    - Back, M. D., Küfner, A. C., & Egloff, B. (2011). "Automatic or the people?" anger on September 11, 2001, and lessons learned for the analysis of large digital data sets. *Psychological Science*, 22(6), 837:838.
    - Frimer, J. A., Aquino, K., Gebauer, J. E., Zhu, L. & Oakes, H. (2015). A decline in prosocial language helps explain public disapproval of the U.S. Congress. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 6591-6594.
    - Wojcik, S., Hovasapian, A., Graham, J., Motyl, M., & Ditto, P. H. (2015). Conservatives report, but liberals display, greater happiness. *Science*, 347, 1243-1246.
    - Frimer, J. A., Brandt, M. J. (under review) Conservatives Display Greater Happiness But Only When They Are in Power: A Linguistic Analysis of the U.S. Congress
    - Dehghani, M., Bang, M, Medin, D., Marin, A., Leddon, E., Waxman, S. (2013). Epistemologies in the Text of Childrens Books: Native and Non-Native Authored Books. *International Journal of Science Education*, 35, 13.
    - *Iliev, R., Axelrod, B., Dehghani, M. (under review) Measuring affect in historical texts: Nation-level happiness in the U.S. is related to affect language in books.

- Week 5: No Class

- Week 6: Text Analysis in Social Sciences – Modern Approaches 1, Project Proposals
    - Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261-266.
    - Johnson, K. M., Garten, J., Dehghani, M. & Graham, J. (in preparation). Advancements in language-based assessments: A practical guide for psychologists
    - Dam, G., & Kaufmann, S. (2008). Computer assessment of interview data using latent semantic analysis. *Behavior Research Methods*, 40(1), 8:20.
    - Sagi, E. & Dehghani, M. (2014). Measuring Moral Rhetoric in Text. *Social Science Computing Review*, 32, 2, 132-144.
    - Diederich, J., Kindermann, J., Leopold, E., & Paass, G. (2003). Authorship attribution with support vector machines. *Applied Intelligence*, 19(1-2), 109:123.
    - Kim, S. M., & Hovy, E. (2006, July). Extracting opinions, opinion holders, and topics expressed in online news media text. *In Proceedings of the Workshop on Sentiment and Subjectivity in Text*. Association for Computational Linguistics.
    - Wienberg, C., Roemmele, M., and Gordon, A. (2013) Content-Based Similarity Measures of Weblog Authors. *The 4th Annual ACM Web Science Conference (WebSci-13*), May 2-4, 2013, Paris, France.
    - Sagae, K., Gordon, A. S., Dehghani, M., Metke, M., Kim, J. S., Gimbel, S. I., Tipper, C., Kaplan, J., Immordino-Yang, M. H. (2013). A Data-Driven Approach for Classification of Subjectivity in Personal Narratives. *Proceedings of the 2013 Workshop on Computational Models of Narrative*.

- *liev, R., Dehghani, M., & Sagi, E. (2014). Automated text analysis in psychology: methods, applications, and future developments. *Language and Cognition*, 1-26.
- *Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, mps028.
- *Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77:101.
- *Diermeier, D., Godbout, J. F., Yu B., & Kaufmann, S. (2011). Language and ideology in congress. *British Journal of Political Science*. 42 (01), 31 – 55.
- *D'Mello, S., & Graesser, A. (2012) Language and discourse are powerful signals of student emotions during tutoring. *IEEE Transactions on Learning Technologies*, 5(4), 304:317.
- *Dehghani, M., Sagae K., Sachdeva, S. & Gratch, J. (2014). Linguistic Analysis of the debate over the Construction of the `Ground Zero Mosque'. *Journal of Information Technology & Politics*, 11, 1-14.

- Week 7: Text Analysis in Social Sciences – Modern Approaches 2, Project Update 1
  - Iliev, R., & Smirnova, A. (2014). Revealing Word Order: Using Serial Position in Binomials to Predict Properties of the Speaker. *Journal of psycholinguistic research*, 1-31.
  - Chae, D. H., Clouston, S., Hatzenbuehler, M. L., Kramer, M. R., Cooper, H. L., Wilson, S. M., ... & Link, B. G. (2015). Association between an Internet-Based Measure of Area Racism and Black Mortality. *PLOS ONE*
  - Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., ... & Seligman, M. E. (2015). Psychological Language on Twitter Predicts County-Level Heart Disease Mortality. *Psychological science*
  - Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... & Seligman, M. E. (2014). Automatic Personality Assessment Through Social Media Language.
  - Schwartz, H. A., Eichstaedt, J., Kern, M. L., Park, G., Sap, M., Stillwell, D., ... & Ungar, L. (2014, June). Towards assessing changes in degree of depression through facebook. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (pp. 118-125).
  - Sap, M., Park, G., Eichstaedt, J., Kern, M., Ungar, L., & Schwartz, H. A. (2014). Developing age and gender predictive lexica over social media. *In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP* (pp. 1146-1151).
  - Garcia, D., & Sikström, S. (2014). The dark side of Facebook: Semantic representations of status updates predict the Dark Triad of personality. *Personality and Individual Differences*, 67, 92-96.

- Week 8: Applications of Network Analysis in Social Sciences, Project Update 1
  - Lerman, K., & Ghosh, R. (2010, May). Information contagion: An empirical study of the spread of news on Digg and Twitter social networks. *In Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM)*.
  - Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*. 110(15), 5802:5805.
  - Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011, October). Our Twitter profiles, our selves: Predicting personality with Twitter. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)* (pp. 180-185). IEEE.
  - Dehghani, M., Johnson, K. M., Sagi, E., Garten, J., Parmar, N. J., Vaisey, S., Iliev, R., & Graham, J. (Under review). Purity homophily in social networks.
  - Na, J., Kosinski, M., & Stillwell, D. J. (2015). When a New Tool Is Introduced in Different Cultural Contexts Individualism–Collectivism and Social Network on Facebook. *Journal of Cross-Cultural Psychology*, 46(3), 355-370.
  - Phan, T. Q., & Airoldi, E. M. (2015). A natural experiment of social network formation and dynamics. *Proceedings of the National Academy of Sciences*, 112(21), 6595-6600.

- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences,* 110(15), 5802-5805.
- Short, M., Mohler, G., Brantingham, P., & Tita, G. (2014). Gang rivalry dynamics via coupled point process networks. *Discrete and Continuous Dynamical Systems - Series B*, 19 (5), 1459-1477
- Hodas, N. O., Kooti, F., & Lerman, K. (2013). Friendship paradox redux: Your friends are more interesting than you. arXiv preprint arXiv:1304.3480.
- *Nettleton, D. F. (2013). Data mining of social networks represented as graphs. *Computer Science Review*, *7*, 1-34.
- *Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3), 75-174.

- Week 9: Audio/Video Analysis and their applications – Manual Analysis
  - Sbarra, D. A, Smith, H. L., & Mehl, M. R. (2012). When leaving your ex, love yourself: Observational ratings of self-compassion predict the course of emotional recovery following marital separation. *Psychological Science*, 23, 261-269.
  - Mehl, M. R., Vazire, S., Holleran, S. E., & Clark, C. S. (2010). Eavesdropping on happiness: Well-being is related to having less small talk and more substantive conversations. *Psychological Science*, 21, 539-541.
  - Ramirez-Esparza, N., Mehl, M. R., Alvarez Bermudez, J., & Pennebaker, J. W. (2009). Are Mexicans more or less sociable than Americans? Insights from a naturalistic observation study. *Journal of Research in Personality*, 43, 1-7.
  - Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, 317, 82.
  - Mehl, M. R. & Robbins, M. L. (2012). Naturalistic observation sampling: The Electronically Activated Recorder (EAR). In M. R. Mehl & T. S. Conner (Eds.), Handbook of research methods for studying daily life. New York, NY: Guilford Press
  - McRae, P. A., Tjaden, K., & Schoonings, B. (2002). Acoustic and perceptual consequences of articulatory rate change in Parkinson disease. *Journal of Speech, Language, and Hearing Research*, *45*(1), 35.
  - White, L., Mattys, S. L., & Wiget, L. (2012). Segmentation cues in conversational speech: robust semantics and fragile phonotactics. *Frontiers in psychology*, *3*.

- Week 10: Audio/Video Analysis and their applications – Automated Analysis
  - Scherer, S., Stratou, G., Gratch, J., & Morency, L. P. (2013). Investigating voice quality as a speaker-independent indicator of depression and ptsd. *Proceedings of Interspeech 2013*.
  - S. Scherer, G. Stratou, M. Mahmoud, J. Boberg, J. Gratch, A. Rizzo, and L.-P. Morency (2013). Automatic behavior descriptors for psychological disorder analysis. In *Proceedings of IEEE Conference on Automatic Face and Gesture Recognition*. IEEE.
  - Scherer, S., Pestian, J., & Morency, L. P. (2013, May). Investigating the speech characteristics of suicidal adolescents. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 709-713). IEEE.
  - Nasir, M., Xia, W., Xiao, B., Baucom, B., Narayanan, S. S., & Georgiou, P. (2015). Still Together?: The Role of Acoustic Features in Predicting Marital Outcome. In *Proceedings of Interspeech*.
  - Gibson, J., Malandrakis, N., Romero, F., Atkins, D. C., & Narayanan, S. (2015) Predicting Therapist Empathy in Motivational Interviews using Language Features Inspired by Psycholinguistic Norms. In *Proceedings of Interspeech*.
  - Bone, D., Lee, C. C., Potamianos, A., & Narayanan, S. (2014). An Investigation of Vocal Arousal Dynamics in Child-Psychologist Interactions using Synchrony Measures and a Conversation-based Model. In *Fifteenth Annual Conference of the International Speech Communication Association.*
  - Ramakrishna, A., Malandrakis, N., Staruk E., & Narayanan, S. (2015) A quantitative analysis of gender differences In movies using psycholinguistic normatives, in: *Proceedings of Empirical Methods in Natural Language Processing* (EMNLP), Lisbon, Portugal.

- Week 11: Project Update 2

- Week 12: Multi Voxel Pattern Analysis (MVPA)
  - Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9), 424–430. doi:10.1016/j.tics.2006.07.005
  - Tong, F., & Pratte, M. S. (2012). Decoding patterns of human brain activity. *Annual Review of Psychology*, 63, 483–509. doi:10.1146/annurev-psych-120710-100412
  - Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191-1195.
  - Just, M. A., Cherkassky, V. L., Aryal, S., & Mitchell, T. M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes. *PloS one*, 5(1), e8622.
  - Hughes, V. (2010). Science in court: head case. *Nature*, *464*(7287), 340-342.
  - Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). " Who" Is Saying" What"? Brain-Based Decoding of Human Voice and Speech. *Science*, *322*(5903), 970-973.
  - Rissman, J., Greely, H. T., & Wagner, A. D. (2010). Detecting individual memories through the neural decoding of memory states and past experience. *Proceedings of the National Academy of Sciences*, *107*(21), 9849-9854.
  - *Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social cognitive and affective neuroscience*, *4*(1), 101-109.
  - *Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*, *45*(1), S199-S209.

- Week 13: Discussion of Deep Learning
  - Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *35*(8), 1798-1828.
  - Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
  - Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems* (pp. 3111-3119).
  - Le, Q. V., & Mikolov, T. (2014). Distributed representations of sentences and documents. arXiv preprint arXiv:1405.4053.
  - Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. *Proceedings of the Empiricial Methods in Natural Language Processing* (EMNLP 2014), 12, 1532-1543.
  - Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013, October). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing* (EMNLP) (Vol. 1631, p. 1642).
  - Mikolov, T., Yih, W. T., & Zweig, G. (2013). Linguistic Regularities in Continuous Space Word Representations. In *HLT-NAACL* (pp. 746-751).
  - *Schmidhuber, J. (2014). Deep Learning in Neural Networks: An Overview. *arXiv preprint arXiv:1404.7828*.

- Week 14 & 15: Final Project Presentations

Statement on Academic Conduct and Support Systems

Academic Conduct

Plagiarism – presenting someone else's ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Section 11, *Behavior Violating University Standards*https://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions/. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct/.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* http://equity.usc.edu/ or to the *Department of Public Safety* http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us. This is important for the safety whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Center for Women and Men* http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage sarc@usc.edu describes reporting options and other resources.

Support Systems

A number of USC's schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* http://dornsife.usc.edu/ali, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information* *http://emergency.usc.edu/*will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.