



AFRL-AFOSR-VA-TR-2018-0350

Recognizing and Modeling Influence in Social Media Language

Yang Liu
UNIVERSITY OF TEXAS AT DALLAS
800 W CAMPBELL RD
RICHARDSON, TX 75080

09/23/2018
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/RTA2

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.					
1. REPORT DATE (DD-MM-YYYY) 08/27/2018		2. REPORT TYPE final		3. DATES COVERED (From - To) Sept 1 2017 - Aug 31 2018	
4. TITLE AND SUBTITLE Recognizing and Modeling Influence in Social Media Language				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER FA9550-15-1-0346	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Liu, Yang				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The University of Texas at Dallas 800 W. Campbell Rd. Richardson, TX 75080				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 875 N Randolph St rm 311 2 suite 325 Arlington, VA 22203				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution A					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Social influence has a profound impact on people's emotion, opinion, and behavior. Prior work has exploited different information, from network analysis and the content. This project aims to focus on the language aspect in addition to social interaction to study influence. We have investigated the following problems in this reporting period: - predict review helpfulness. The review usefulness used in previous work has not been systematically tested. Second, whether a user rates a review as helpful depends on not only the review content but also the context in which it was written, such as whether the user agrees with the reviewer's opinion. Merely using features taken from the review limits a system's predictive power. Our work so far has focused on identifying these key research issues in existing work. - extract implicit aspects in reviews. So far we have found some characteristics of implicit aspects (they are common, overlap with explicit aspects, functionality of opinions, attached to specific attributes) and just started data collection effort.					
15. SUBJECT TERMS social influence, review, opinion, aspects					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Dina Caplinger
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code) 972-883-2312

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Recognizing and Modeling Influence in Social Media Language

PI: Yang Liu

The University of Texas at Dallas

AFOSR Award: FA9550-15-1-0346

Final report

2017/09-2018/08

Social influence has a profound impact on people's emotion, opinion, and behavior. This project aims to focus on the language aspect in addition to social interaction to study influence and related aspects in social media language. In particular, in the past year our objective is advancing performance for predicting review helpfulness and extracting implicit aspects in social media text.

1) Review Helpfulness Prediction

Product reviews written by users in online stores, such as Amazon.com, have become a key resource for consumers looking to make well-informed decisions. Given that the quality of these reviews can vary greatly and that products frequently have too many reviews for a single customer to read, there is an increasing need for systems that can automatically determine the quality of product reviews. Researchers have used user votes such as those elicited by Amazon's "Did you find this review helpful?" (Yes/No) prompt, as indicators of the "helpfulness" of a review. The ratio of "helpful" votes a review receives has been taken as a measure of its helpfulness. The goal of helpfulness prediction is to build systems that predict this ratio.

Results: We noticed that existing work on helpfulness prediction has several key weaknesses, which we attempt to address. First, existing work has focused on finding useful features from a given review for helpfulness prediction, but their usefulness have not been systematically tested. Second, whether a user rates a review as helpful or not depends on not only the review content but also the context in which it was written, such as whether the user agrees with the reviewer's opinion. Merely using features taken from the review limits a system's predictive power. Finally, systems were evaluated on different datasets, making it hard to determine the state of the art. Our work so far has focused on identifying these key research issues. Our results are available in the following publication:

Modeling and Prediction of Online Product Review Helpfulness. Gerardo Ocampo Diaz and Vincent Ng. Proceedings of the 56th Annual Meeting of the ACL (Vol. 1: Long Papers), 2018.

2) Implicit Aspect Extraction

User reviews in online stores and review websites are gold mines of information for manufacturers, retailers, and consumers. Reviewers' opinions are embedded in text reviews, but the large number of reviews posted every day makes it infeasible to manually process them. The goal of aspect extraction is to automatically identify opinions expressed on different aspects of a product. The two main types of aspect extraction are: a) explicit aspect extraction, which attempts to recover expressed opinions on explicit aspects of a product; and b) implicit aspect extraction, which attempts to recover opinions expressed implicitly. Most work has focused on explicit aspect extraction. Implicit aspect extraction is arguably harder and much-less studied.

Worse still, there is no consensus on how implicit aspect extraction should be performed and evaluated.

Since there are no public datasets for implicit aspect extraction, we attempt to build the first such dataset, which should allow us to a) encourage researchers to pursue implicit aspect extraction, and b) define what implicit aspect extraction should consist of and how it should be evaluated.

Results: We do not have publishable results yet, but our research so far has yielded the following findings. First, implicit aspects are quite common in reviews: they are at least as frequent as explicit aspects. Second, implicit and explicit aspects do overlap, i.e., aspects that are mentioned explicitly are sometimes present implicitly. Third, a lot of the difficulties found when performing implicit aspect extraction seem to stem from the fact that aspect based sentiment analysis makes the assumption that opinions are always expressed in terms of concrete aspects of a product, where in reality, opinions are many times expressed in terms of functionality, i.e., product behavior. Finally, inferring specific attributes is sometimes not enough. e.g., inferring resolution from "Images are very detailed" is not enough, extra steps must be taken to infer the specific component on which opinions are expressed (in this case, "screen"). This has also led us to reconsider the definition of "aspect".