Learning from Personal Longitudinal Dialog Data

Charlie Welch (cfwelch@umich.edu), Verónica Pérez-Rosas (vrncapr@umich.edu),
Jonathan K. Kummerfeld (jkummerf@umich.edu), Rada Mihalcea (mihalcea@umich.edu)
University of Michigan, Ann Arbor

Motivation

Large amounts of personal data exist on social media platforms

These can be used for dialog related tasks.

1. Common Utterance Prediction – classifying what someone will say next from a common set of utterances (e.g. Allo, Inbox auto-response).
2. Response Time Prediction – determining when a system should reply or to predict when someone will reply to you.
3. User Attribute Prediction – determining the relationship between speakers to better understand personal interactions.

What is in one person's data?

Can we perform these predictions?

Will our findings generalize to others?

Data and Annotation

This table contains statistics about our data from a single individual.

	Author	Others	All
Total Msgs	237,300	216,766	454,066
Unique Msgs	165.536	168,041	326,243
Total Tokens	1,370,916	1,602,607	2,973,523
Unique Tokens	38,937	48,005	68,985
Avg Tokens / Msg	5.78	7.39	6.55

We annotate 104 conversation partners with over 100 messages exchanged for 7 attributes:

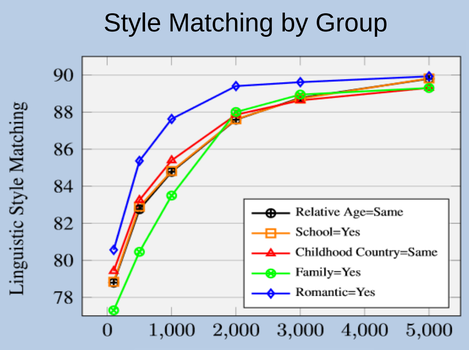
- Family
- Romantic Relationship
- Relative Age
- Childhood Country
- Gender
- Known from School
- Known from Work

For prediction tasks we use a context window of five messages. Two examples are shown to the right.

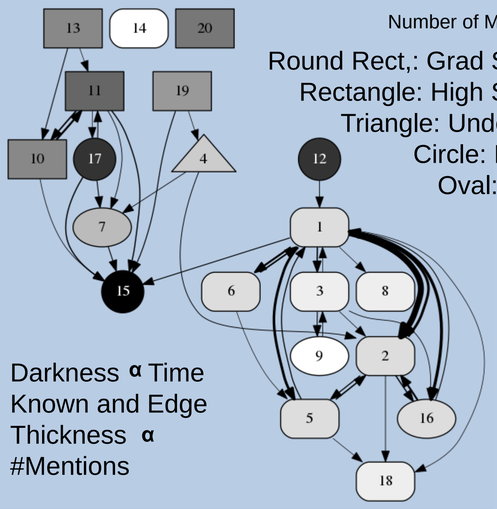
MsgNum	Time	Message
c1msg1	15:45:06	P: Wanna go grab some coffee?
c1msg2	15:45:20	A: yeah
c1msg3	15:45:25	P: Sweet!!!!
c1msg4	15:45:29	P: Meet in the lobby?
c1msg5	15:45:52	A: okay
c2msg1	12:21:00	P: Perfect!!
c2msg2	15:56:22	P: Want to go get Thai?
c2msg3	16:01:18	P: I'll take it you're sleeping lol
c2msg4	16:19:59	A: Yeah
c2msg5	16:20:08	A: I mean yeah I was sleeping

Longitudinal Analyses

Style matching is similarity of use of 8 types of function words



Interaction Groups
Nodes are people
Edges are mentions

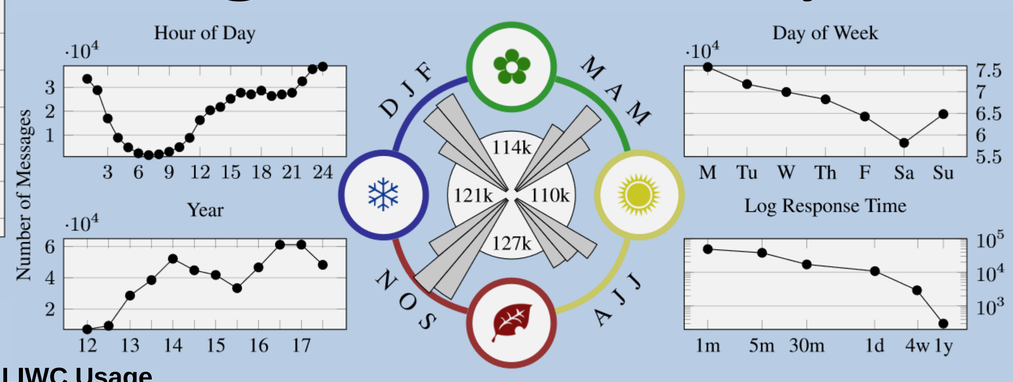


Round Rect,: Grad School
Rectangle: High School
Triangle: Undergrad
Circle: Family
Oval: Other

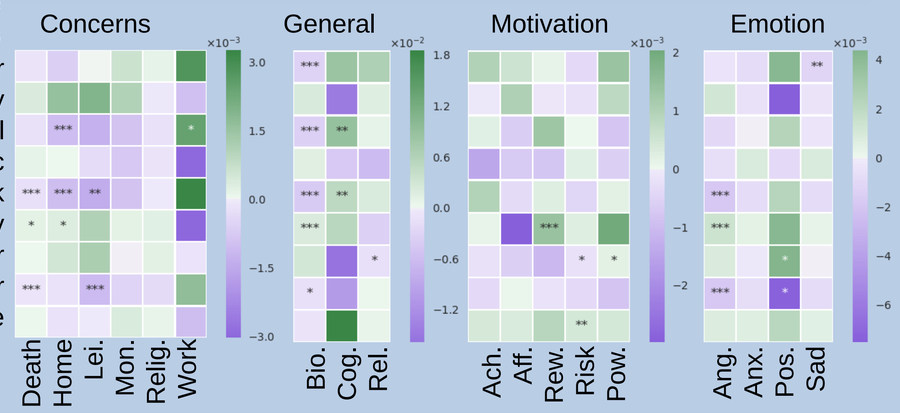
Darkness ∝ Time Known and Edge Thickness ∝ #Mentions

Group Attributes – value calculated as: in - out

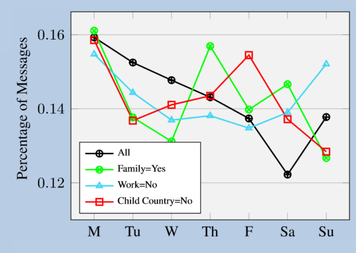
* p < 0.10
** p < 0.05
*** p < 0.01



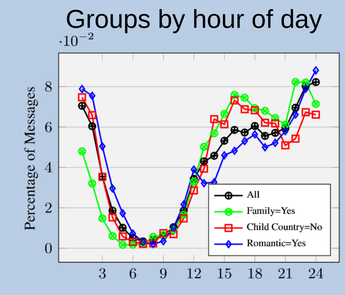
LIWC Usage Differences



Groups by day of week



Groups by hour of day



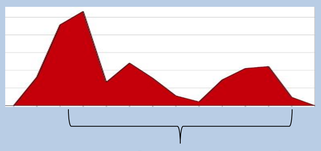
LDA Topic	Sample Words
Communication	phone, email, send, number, text, message
Positive	lol, haha, yeah, like, good, just, nice, pretty
Negative	don't, know, feel, people, bad, weird, mean
Data	data, right, think, know, mean, words
Entertainment	game, play, picture, video, buffy, buy
Food & Time	time, home, want, eat, work, today, food
Planning	go, come, tomorrow, party, tonight, home
Begin & End	hi, hey, sup, yo, thanks, sleep, bed, tired
Sharing Media	youtube, like, yeah, good, music, good, site
Read & Write	read, paper, write, learning, book, language
Travel	pizza, food, airport, russian, english, mom
Computing	know, code, file, just, check, add, run, make
Living & Money	`current_city', car, buy, live, pay, house

Model

Time features: Month, day, year, season, time, difference in times, e.g. 14 seconds

Attributes are used as features e.g. {female, older, co-worker, etc.}

Time	Message
15:45:06	P: Want to grab some coffee
15:45:20	A: yeah
15:45:25	P: Sweet!!!!
15:45:29	P: Meet in the lobby?
15:45:52	A: okay

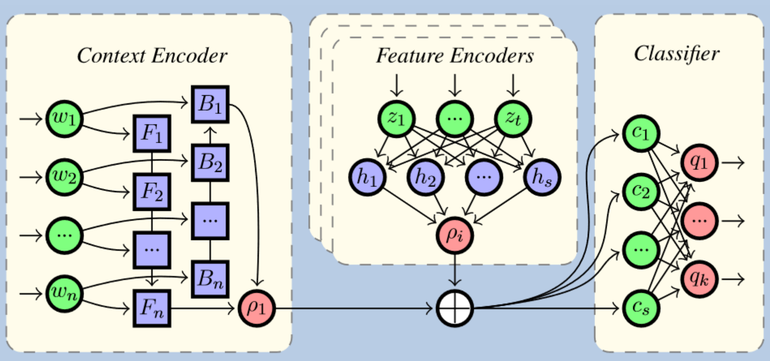


Message frequency taken from distribution of messages over time

GloVe Common Crawl Embeddings input to a BiLSTM

LIWC Vectors + Similarity between usage for a subset gives style matching

want to … coffee



Next Message: yes, haha, okay, oh, nice, <other>

Response Time: <90sec, <10min, <1day, >1day

Speaker Attributes: Predict all 7

Results

Results for next message prediction and response time with feature ablations. Seven volunteers ran experiments on their data

Model	Next Message	Response Time
Single Person		
Majority Class	34.0	61.9
MsgEmb	46.5	64.4
MsgEmb + Time	47.0	67.5
MsgEmb + LIWC	47.6	64.4
MsgEmb + StyleMatch	46.7	64.4
MsgEmb + MsgFreq.	47.6	64.8
MsgEmb + Attributes	46.9	64.5
All Features	50.0	68.0
Seven Volunteers		
Majority Class	32.2±9.6	66.4±9.8
MsgEmb	41.2±6.2	69.8±7.4
All Features	45.0±7.2	74±5.6

Speaker attribute models have 3 variants:

1. Single attribute decoders
2. Joint decoders
3. Single attribute decoders + attribute features

Context window averages are in parentheses next to the person-level majority vote

Attributes	Baseline LSTM	Best Models
Family	94.2 (92.0)	94.2 (93.9)
Romantic Relationship	91.3 (86.0)	91.3 (90.9)
Relative Age	45.2 (39.2)	51.9 (47.2)
Childhood Country	79.8 (75.7)	88.5 (85.7)
Gender	86.5 (63.7)	86.5 (68.6)
Classmate	73.1 (64.6)	78.8 (74.3)
Co-worker	80.8 (69.5)	87.5 (83.4)

Acknowledgments: This material is based in part upon work supported by IBM under contract 4915012629 and by the Michigan Institute for Data Science. Any opinions, findings, conclusions or recommendations expressed above are those of the authors and do not necessarily reflect the views of IBM or the Michigan Institute for Data Science.