In search of emotional traces

Detection of depression through social media

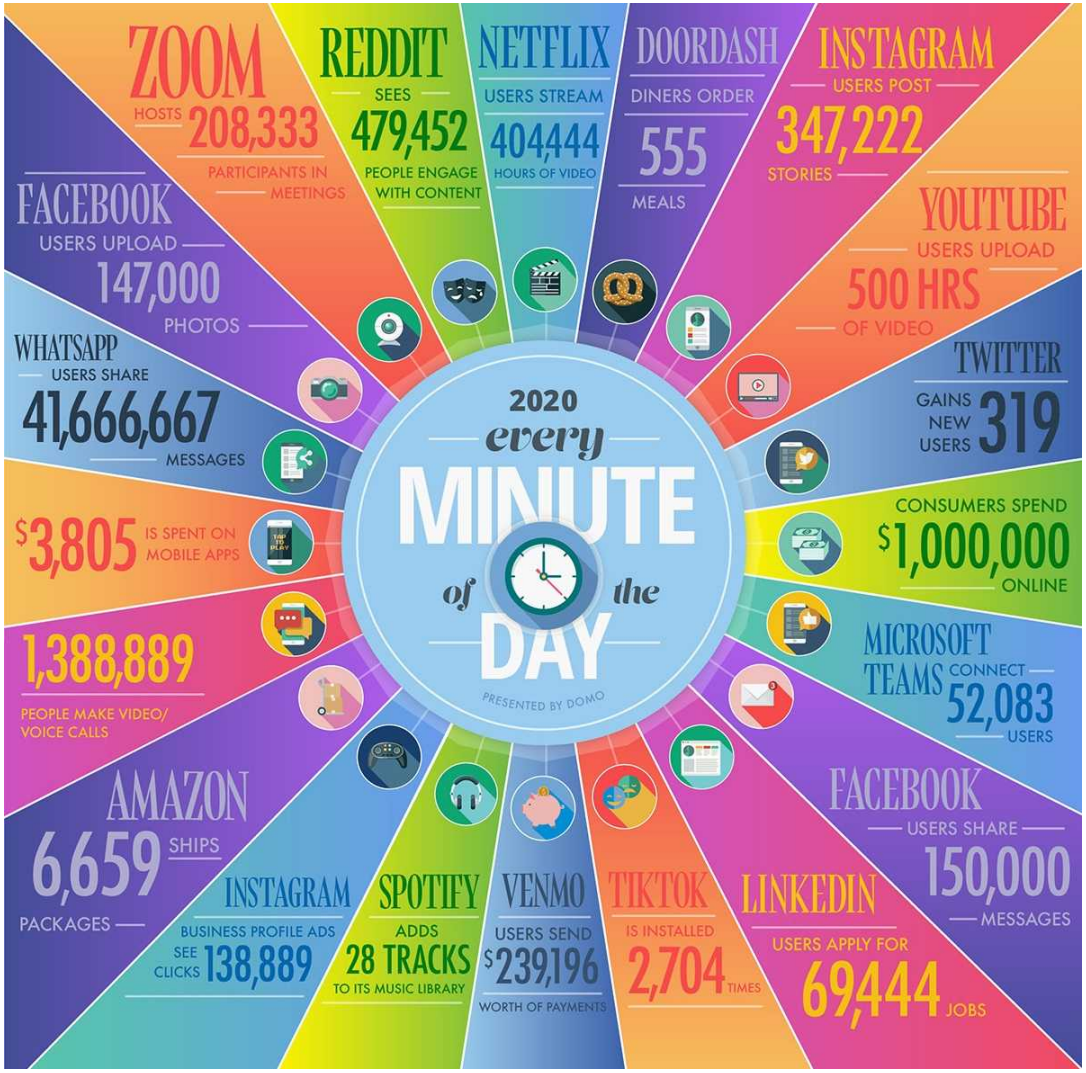
Manuel Montes y Gómez

Laboratorio de Tecnologías del Lenguaje
Instituto Nacional de Astrofísica, Óptica y Electrónica, México
http://ccc.inaoep.mx/~mmontesg/
mmontesg@inaoep.mx

2020 IEEE Autumn Meeting on Power, Electronics and Computing, ROPEC,
November 6, 2020

A day in a life

Here, there and everywhere



https://www.visualcapitalist.com/every-minute-internet-2020/

Some questions to think about:

- What can we learn from all this data?
- What can we extract or infer from the users?

I am looking through you

My own example:

- Male
- Computer Science
- I have (at least) one son
- I like the Beatles
- I like classical music
- I speak Spanish and English
- I live in Puebla
- …
- I will be at ROPEC 2020

Author profiling

- It consists in knowing as much as possible about an author, just by analyzing his/her texts.
  - Age, gender, social/economic status, level of studies, occupation, nationality, religion, etc.

Author → *user* of social media
Text → *multimedia* content

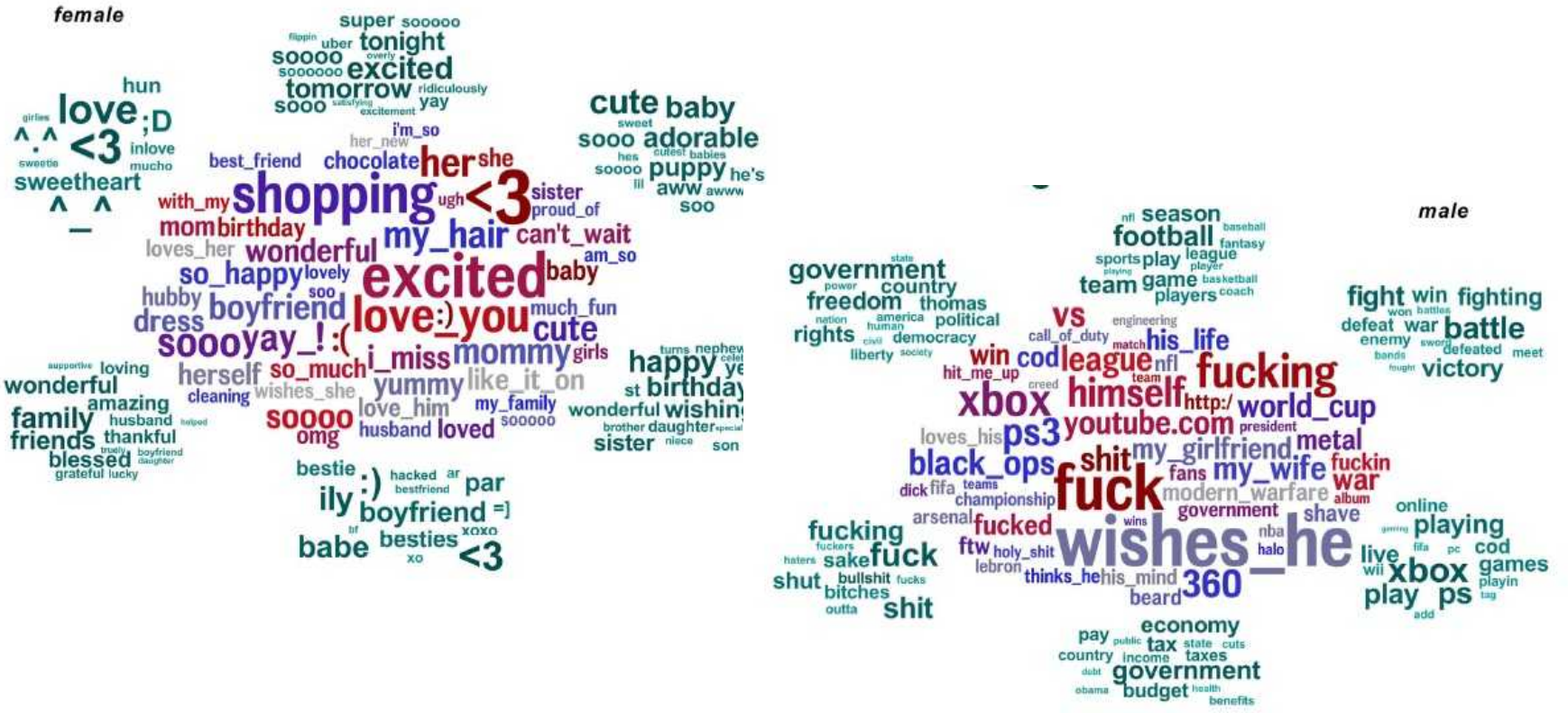
- AP has been mainly used for:
  - Marketing
  - Security purposes

Author profiling – main approaches

- Approached as a *text classification task*
  - where profiles represent the classes to discriminate.
- Focus on determining relevant features
  - Content features (*words*, n-grams, topics)
  - Style features (punctuation marks, emoticons, stopwords, POS tags, etc.)
- Recently, CNNs and RNNs have shown slightly better results than traditional ML approaches

- Overview of the 7th Author Profiling Task at PAN 2019: Bots and Gender Profiling. Francisco Rangel and Paolo Rosso. Notebook Papers, **CLEF 2019**.
- **Overview of the 6th Author Profiling Task at PAN 2018: Cross-domain Authorship Attribution and Style Change Detection**. Francisco Rangel, Manuel Montes-y-Gómez, Martin Potthast, and Benno Stein. Working Notes Papers, **CLEF 2018.**

Frequent words by women and men



Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. Schwartz et al. PLoS ONE 8(9), 2013.

Topics from Mexican Twitter users

Male:

Female:

Overview of MEX-A3T at IberEval 2018: Authorship and Aggressiveness Analysis in Mexican Spanish Tweets. Miguel Á. Álvarez-Carmona et al. **IberEval 2018**.

Moving a step forward

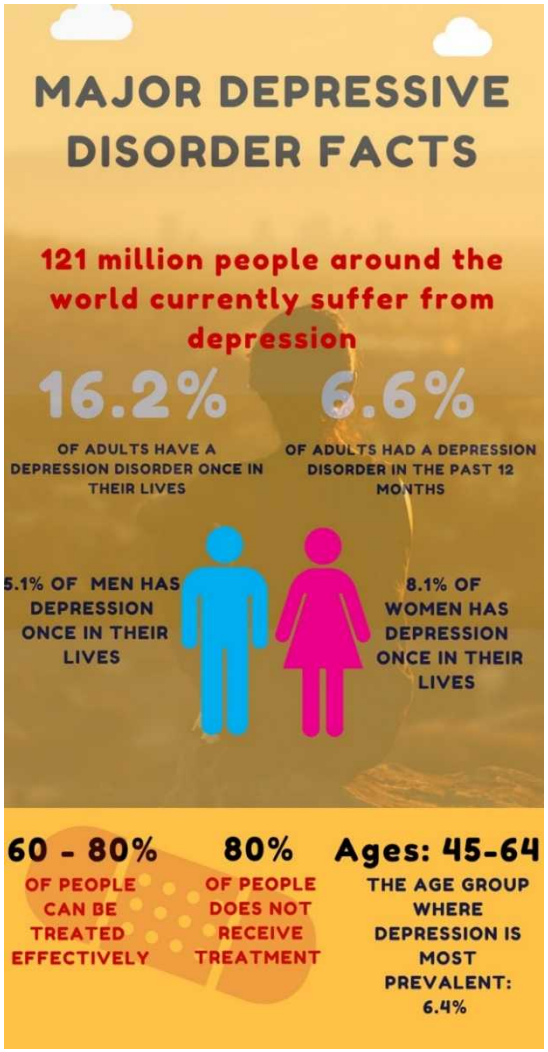
- Not only predict sociodemographic attributes from the users, but also do a kind of *behavior profiling*.
- The goal is to predict some complex (behavioral) characteristics of users such as:
  - Their personality traits
  - If they suffer from a mental illness, e.g., *depression*
- Also to detect special kinds of users:
  - Bots and spammers
  - Aggressive users
  - Sexual predators

Outline of the talk

- Predicting depression via social media
  - Introduction to the task
- Three methods based on the analysis of emotions:
  - The **B**ag **o**f **S**ub-**E**motions (BoSE) representation
  - Capturing static and dynamic emotional patterns
  - A deep model based on the BoSE representation
- Our ongoing and future work

Some depression facts

- Psychological syndrome associated with distress and disability.
- It affects the thinking, feeling, mood, and behavior.
- According to the OMS:
  - 800,000 people commit suicide every year as a result of depression
  - It is the second cause of dead among people from 15 to 29.
  - In many countries more than 75% of the persons suffering depression *are not treated*



https://www.who.int/news-room/fact-sheets/detail/depression
https://barendspsychology.com/online-therapy-mental-disorders/

Depression in Mexico

- Of the health budget in Mexico:
  - Only about 2% is allocated to mental health.
  - The World Health Organization recommends 5 to 10%.
  - it lacks a focus on *prevention* rather than containment.

https://eptanetwork.org/database/policy-briefs-reports/1543-mental-health-in-mexico-incytu
https://www.eluniversal.com.mx/english/mexico-ignores-mental-health
https://www.mientrastantoenmexico.mx/mexico-la-depresion-genera-perdidas-mas-14-billones-dolares

Anorexia in Mexico

- Each year there are 20 thousand new cases of anorexia and bulimia.
- 10% of sick young people with anorexia and 17% sick with bulimia tried to *commit suicide*
- Eating disorders seem to focus only on people's body image, food and weight.
  - However, in many cases they are related to other issues depression!

Encuesta Nacional de Salud y Nutrición 2018. Secretaria de Salud, Gobierno de México. https://ensanut.insp.mx/encuestas/ensanut2018/doctos/informes/ensanut_2018_presentacion_resultados.pdf.
http://insteractua.ins.gob.pe/2018/02/anorexia-infografia.html

A new opportunity for their detection

- Depression alters the way people move and sleep, but it also alters the way they *speak and write*.
- Now social media is one of the most common ways people communicate each other.
- Social media analysis *increases the chances to detect* people that present signs of depression/anorexia, and to guide them to professional help as soon as possible.

- The Secret Life of Pronouns: What Our Words Say About Us. J. Pennebaker. Bloomsbury USA, 2011.
- Psychological aspects of natural language use: Our words, ourselves. J. Pennebake et al. Annual review of psychology 54 (2003).
- Me, myself, and i: self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. T. Brockmeyer et al. Frontiers in Psychology 6 (2015).

eRisk shared task

- A forum dedicated to early risk prediction on the Internet, organized since 2017, at the *CLEF conference*.
- Focus on topics related to health and safety; particularly the detection of *depression* and *anorexia*.
- Data from Reddit; positive instances are users that explicitly mentioned they have been diagnosed with depression.

Collection	Training		Test	
	Positive	Negative	Positive	Negative
Dep2017	83	403	52	349
Dep2018	135	752	79	741
Anx2018	20	132	41	279
Anx2019	61	411	73	742

https://erisk.irlab.org/

A Test Collection for Research on Depression and Language Use. D. E. Losada, F. Crestani**. CLEF 2016.**

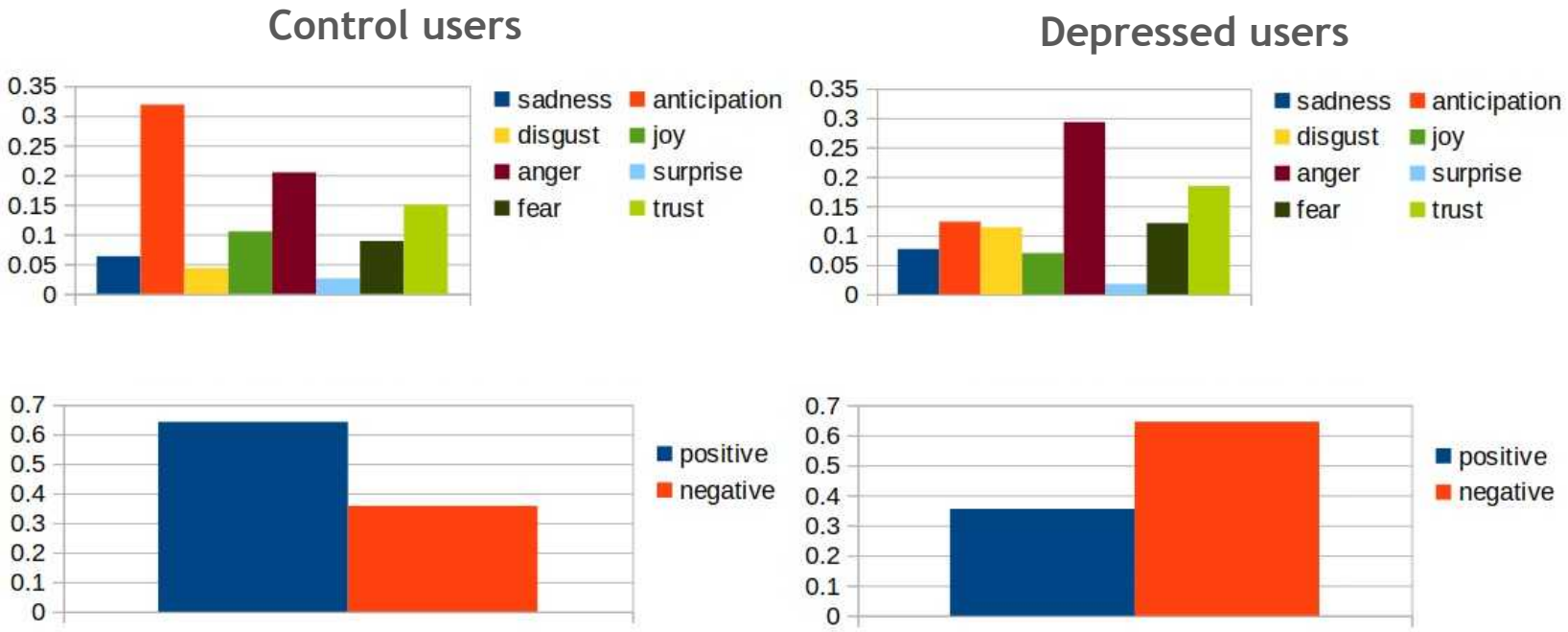
Idea 1: an emotion-based representation

- Depression is a serious mental illness that can interfere with persons' lifes.
  - Usually alters their interests, activities and feelings.
- Some *emotions* are particularly related to depression, for example, fear, anger and shame.
- *Negative words* are also common in persons suffering from depression.

Hypotesis: *control and depressed social media users must show different emotion distributions in their posts*.

What about mood swings? identifying depression on twitter with temporal measures of emotions. C. Xuetong, D. Martin, W. Thomas, and E. Suzanne. **WWW 2018**.

Distribution of emotions at eRisk posts



- We used a lexical resource which contains 5,000 words per emotion and around 12,000 words per sentiment.

Crowdsourcing a word-emotion association lexicon. Saif M. Mohammad and Peter D. Turney. Computational Intelligence, 2013.

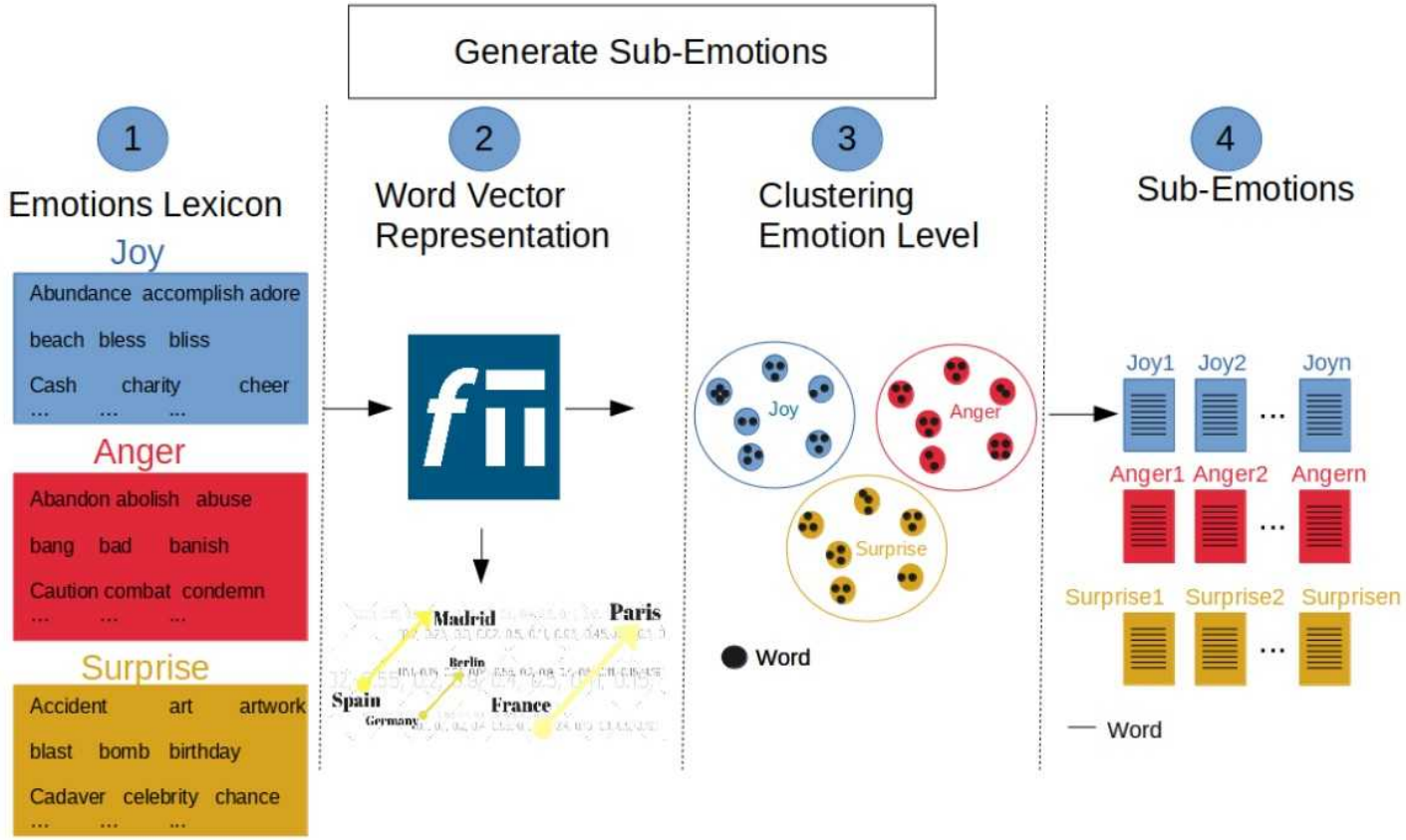
Motivation of our idea

- Emotions are related with many different things.
  - For example, surprise could be related with an accident but also with a party.

- Our idea is to *model emotions in a fine-grained way*, and use them to build a new more detailed representation.

Using fine-grained emotions (1)

The idea is to represent users by an histogram of -automatically discovered- fine-grained emotions



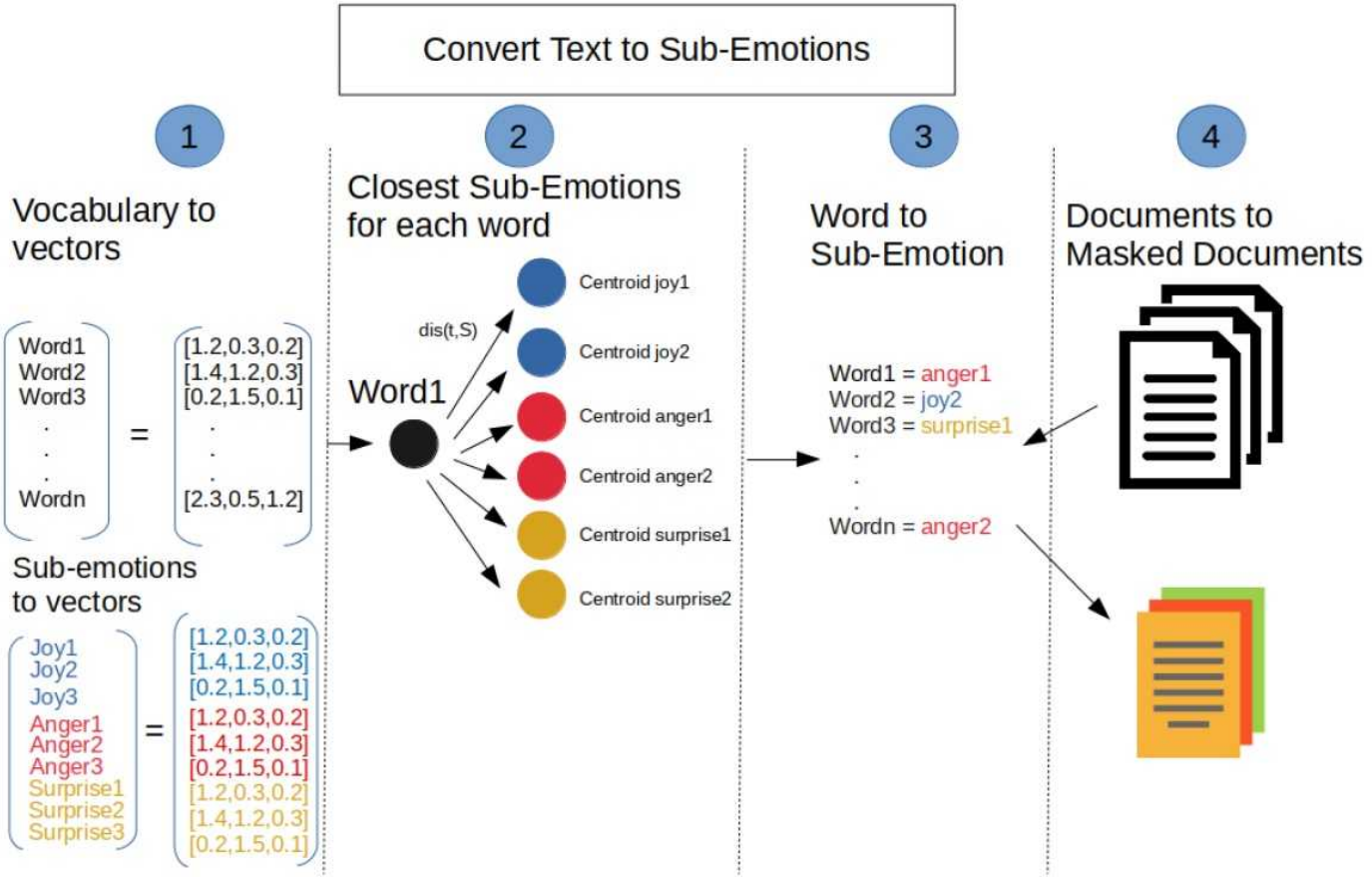
Detecting Metal Health Disorders in Social Media using Fine-Grained Emotions. Ezra Aragón, Pastor López-Monroy, Manuel Montesy-Gómez, Luis Carlos González. **NAACL-HTL 2019.**

Examples of our fine-grained emotions

Anger			Joy		
abomination	growl	battle	accomplish	bounty	charity
fiend	growling	combat	achieve	cash	foundation
inhuman	thundering	fight	gain	money	trust
abominable	snarl	battler	reach	reward	humanitarian
unholy	snort	fists	goal	wealth	charitable
Surprise			**Disgust**		
accident	art	magician	accusation	criminal	cholera
crash	museum	wizard	suspicion	homicide	epidemic
disaster	artwork	magician	complaint	delinquency	malaria
incident	gallery	illusionist	accuse	crime	aids
collision	visual	sorcerer	slander	enforcement	polio

Using fine-grained emotions (2)

The idea is to represent users by an histogram of -automatically discovered- fine-grained emotions



Results of our approach

F1 results over the positive class

Method	Dep'17	Dep'18	Anor'18
BoW-unigrams	0.60	0.58	0.69
BoE-unigrams	0.57	0.60	0.50
BoSE-unigrams	0.61	0.61	**0.82**
BoW-ngrams	0.59	0.60	0.69
BoE-ngrams	0.61	0.58	0.58
BoSE-ngrams	**0.64**	**0.63**	0.81
BiLSTM-Glove	0.39	0.46	0.46
BiLSTM-word2vec	0.53	0.48	0.56
CNN-Glove	0.45	0.51	0.54
CNN-word2vec	0.55	0.48	0.57

Detecting Metal Health Disorders in Social Media using Fine-Grained Emotions. Ezra Aragón, Pastor López-Monroy, Manuel Montesy-Gómez, Luis Carlos González. **NAACL-HTL 2019.**

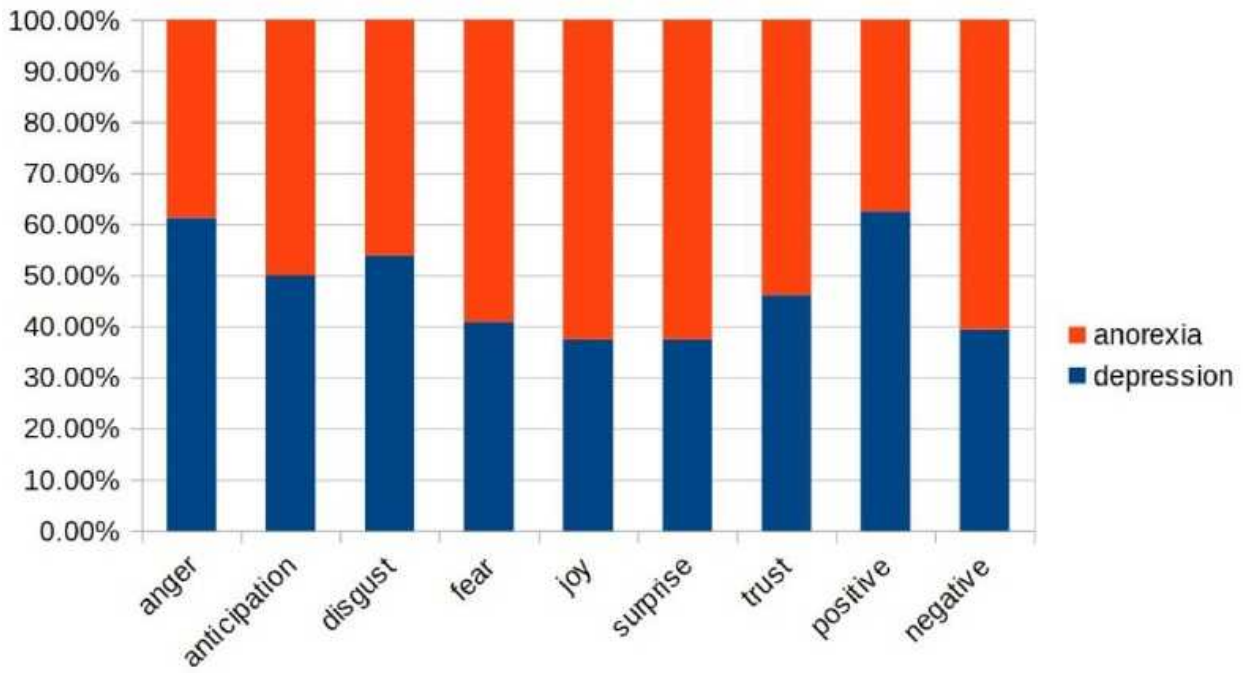
Relevant fine-grained emotions

Depression	
anger1	abandoned, deserted, unattended
anger11	unsociable, crowd, mischievous
anticip10	disappointed, inequality, infidelity
disgust16	unsatisfactory, dilution, influence
disgust11	insecurity, desolation, incursion
fear17	hysterical, immaturity, injury

Anorexia	
anger4	bruising, contusion, bleeding, fracture
disgust32	breakdown, fight, crushed, abandoned
disgust21	stomach, intestinal, bile, esophagus
negative65	bathroom, toilet, washroom
anticip10	hurting, refused, anxious, afraid
anticip12	ashamed, embarrass, upset, disgust
fear19	food, eating, eat, consume

- Depression:
  - Mainly *negative topics*
  - *Anger* subemotions are the most frequent, related to dilution, insecurity, and desolation.
- Anorexia:
  - Negative topics and disgust subemotions are the most frequent
  - Several subemotions related to embarrassment, fear, shame; *self-harm and eating topics*.

Depression and anorexia emotions



- Distribution of the emotions, as captured from the 100 most frequent sub-emotions.
- Depression and anorexia have their *own emotional silhouette*!

Detecting mental disorders in social media through emotional patterns. Mario Ezra Aragón, A.Pastor López-Monroy, Luis C. González-Gurrola, Manuel Montes-y-Gómez. **Submitted to Applied Intelligence**.

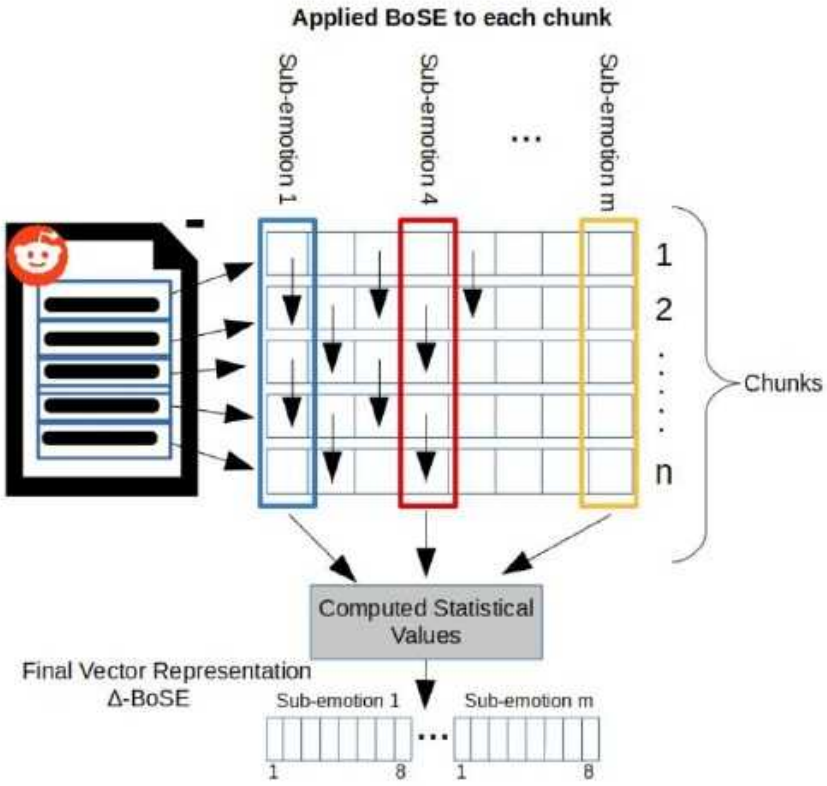
Idea 2: capturing the emotional variability

- Previous hypothesis: people with depression and anorexia show different types of emotions (than control users).
  - We represented users by an histogram of sub-emotions
- New hypothesis: people with these mental disorders tend to expose greater *emotional variability* than a healthy person.
  - The new idea is to represent each user by a set of statistical values that describe the frequency changes of the sub-emotions over time

Δ-BoSE: a dynamic sub-emotion representation

It aims to capture *temporal emotional patterns*:

1. We divide the post history of each user in *n* parts or chunks.
2. We calculate the BoSE representation of each chunk
3. We represent each subemotion by a Δ-vector of eight statistical values: mean, sum, max, min, std, variance, median.



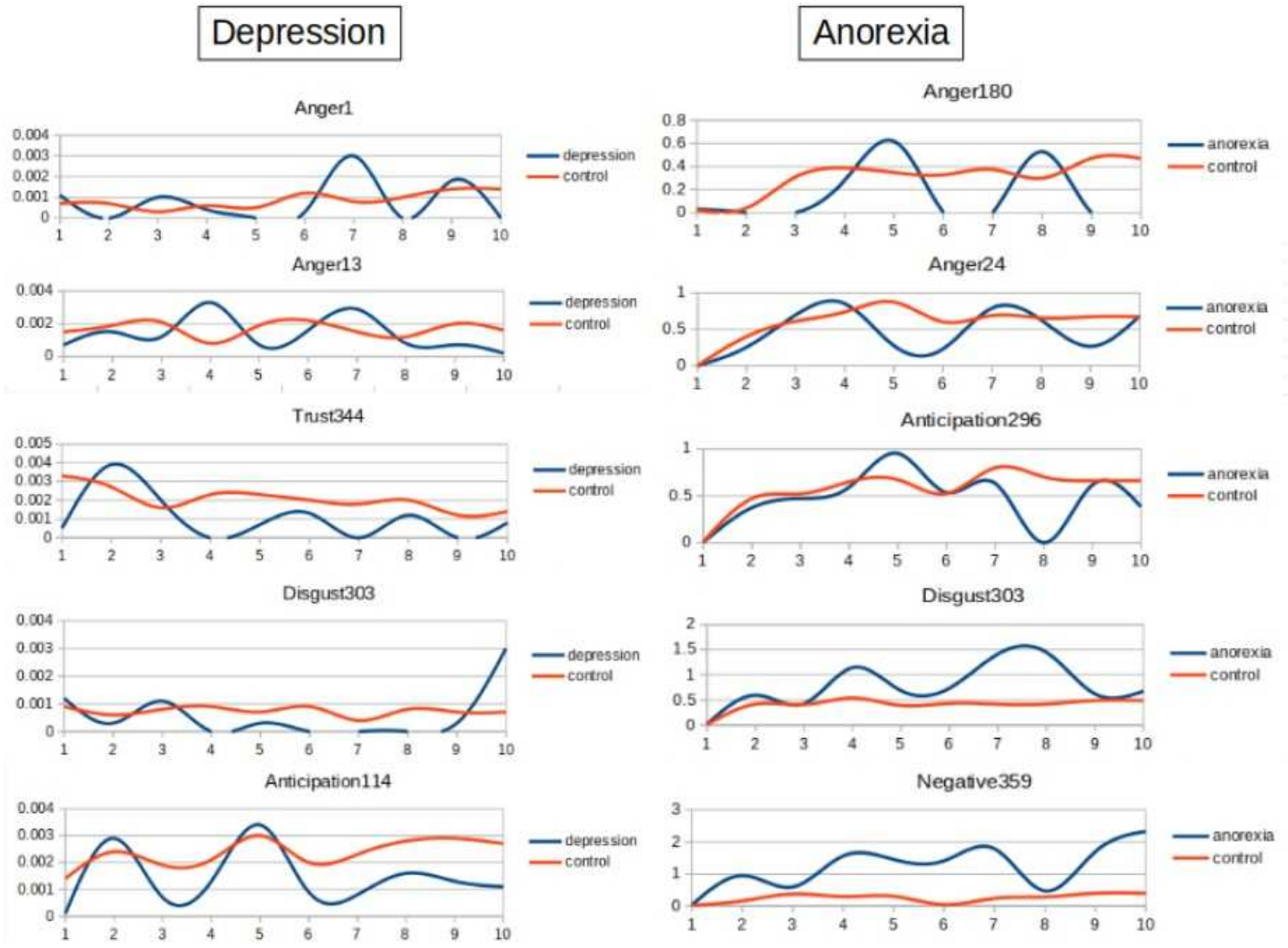
Detecting mental disorders in social media through emotional patterns. Mario Ezra Aragón, A.Pastor López-Monroy, Luis C. González-Gurrola, Manuel Montes-y-Gómez. **Submitted to Applied Intelligence**.

Results of the new approach

	Depression'18	Anorexia'18
BoSE	0.63	0.82
Δ-BoSE	0.53	0.79
Early Fusion	0.62	0.77
Late Fusion	**0.64**	**0.84**

- Using only BoSE is more informative than only capturing sub-emotion changes over time.
- Nonetheless, when we *fusion both representations*, the performance of the classification was improved.
  - Best results using the **late fusion approach**, where we combined the decision of two classifiers by means of an OR gate over their corresponding decisions.

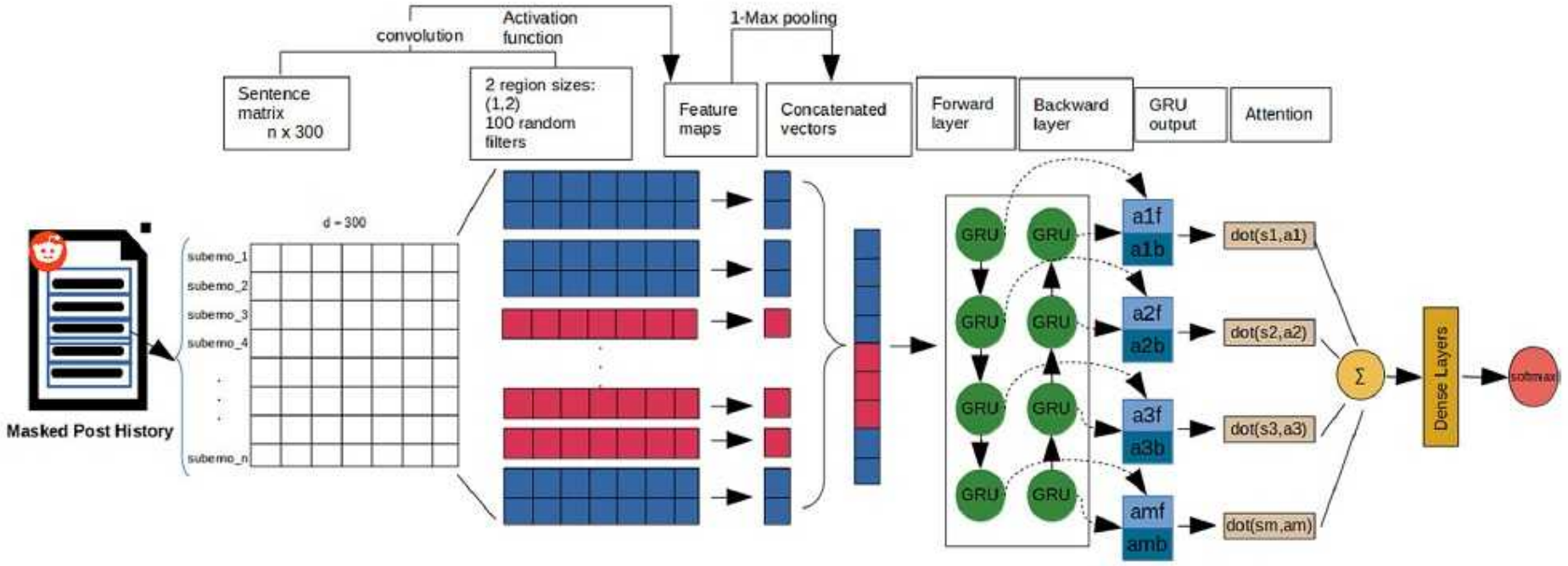
What is captured by Δ-BoSE?



Idea 3: attention to emotions

- This method extends our preceding work by evaluating a deep learning architecture.
  - It consists on processing the users' posts masked with sub-emotions using a deep learning model with an attention mechanism.
- The intuition is that *not all sub-emotions contribute equally* to the representation of the users.
  - Using the attention mechanism, we can extract the (sequences of) sub-emotions that are most important to the meaning of each post and aggregate this importance to the representation.

Deep emotion attention model



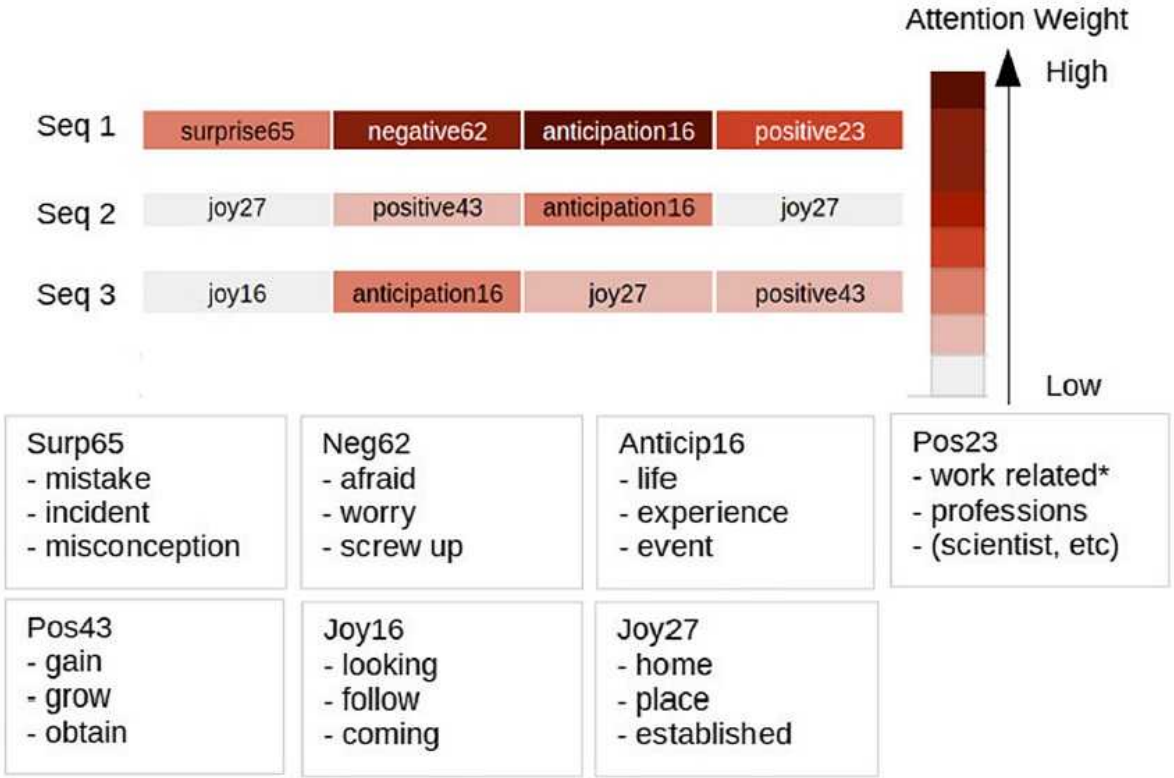
Attention to Emotions: Detecting Mental Disorders in Social Media. Mario Ezra Aragón, A.Pastor López-Monroy, Luis C. González-Gurrola, Manuel Montes-y-Gómez. **TDS 2020.**

Preliminary results

	Method	Anorexia	Depression
Baselines	BoW	0.67	0.54
	N-grams	0.66	0.54
	RNN-Glove	0.65	0.46
	RNN-word2vec	0.65	0.48
	CNN-Glove	0.67	0.51
	CNN-word2vec	0.66	0.48
	Deep-Attention (with FastText)	0.66	0.50
eRisk results	First place	**0.71**	**0.64**
	Second place	0.68	0.60
	Third place	0.68	0.58
Our methods	Deep Emotion Attention (broad emotions)	0.71	0.46
	Deep Emotion Attention (sub-emotions)	**0.79**	**0.58**

Data set	Training		Test	
	NC	C	NC	C
dep eRisk'18	135	752	79	741
anor eRisk'19	61	411	73	742

What is captured by the attention model?



- Greater attention to “anticipation16” (related to life, experience, and events) when it is close to sub-emotions related to worries, afraid, mistakes, and incidents.
- Lower attention to “anticipation16” when it is close to gain, growth, home, and place.

Our works, a team effort

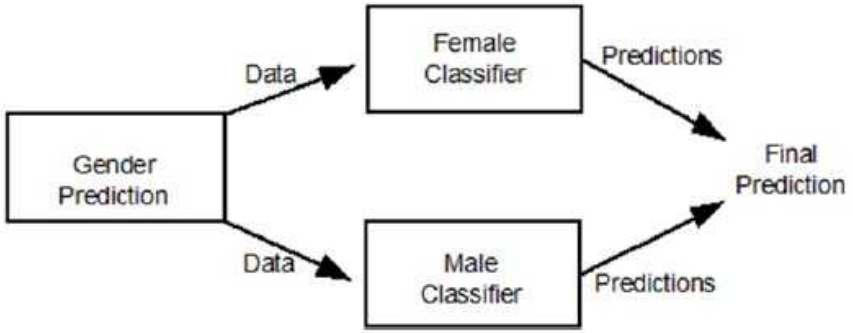
- *Detecting Metal Health Disorders in Social Media using Fine-Grained Emotions*. Mario Ezra Aragón, Pastor López-Monroy, Manuel Montes-y-Gómez, Luis Carlos González. **NAACL-HLT 2019**.
- *An Incremental Framework for Simple and Eective Early Depression Detection*. Sergio G. Burdisso, Marcelo Errecalde, Manuel Montes-y-Gomez. **Expert Systems With Applications**, Vol. 133, pp. 182–197, 2019.
- *Crosslingual Depression Detectionin Twitter using Bilingual Word Alignment*. Laritza Coello, Rosa María Ortega, Manuel Montes, Luis Villaseñor. **CLEF 2019**.
- *Attention to Emotions: Detecting Mental Disorders in Social Media*. Mario Ezra Aragón, A. Pastor López-Monroy, Luis C. González-Gurrola, Manuel Montes-y-Gómez. **TSD 2020**.
- *t-SS3: a Text Classifier with Dynamic N-Grams for Early Risk Detection over Text Streams*. Sergio Burdisso, Marcelo Errecalde, Manuel Montes-y-Gómez. **Pattern Recognition Letters**. Volume 138, October 2020.
- *Revealing traces of Depression through Personal Statements Analysis in Social Media*. Rosa Ortega-Mendoza, Delia Irazú Hernández, Manuel Montes, Luis Villaseñor-Pineda. **Submitted to the Journal of Artificial Intelligence in Medicine**.
- *Detecting Mental Disorders in Social Media Through Emotional Patterns - The case of Anorexia and Depression.* Mario Ezra Aragón, Adrián Pastor Lopez-Monroy, Luis Carlos González-Gurrola and Manuel Montes-y-Gómez. **Submitted to IEEE Transactions on Affective Computing**.
- *Attracted to unstable people: mental disorders detection as a one-class classification problem.* Juan Aguilera, Delua Irazú Hernández, Rosa María Ortega-Mendoza, Manuel Montes-y-Gómez. **Submitted to Applied Intelligence.**

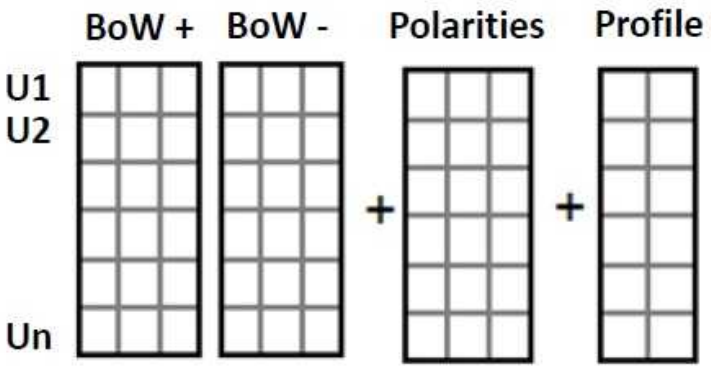
The long and winding road

- **Deep learning models** for early detection of depression
- Synergy between **author profiling** and the detection of depression
- **Crosslingual depression** detection: using English resources to analyze Mexican users

Not all are equal

- The way people express themselves is related to their *profile characteristics*.
- Depression detection could be improved by considering their gender and age, as well as by jointly modeling topics and emotions.

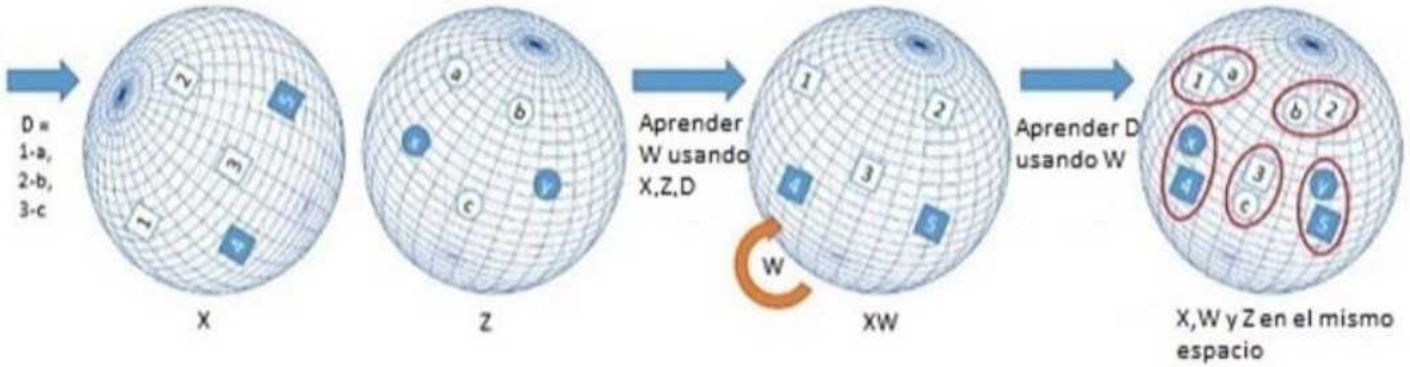




José de Jesús Titla Tlatelpa. **Detección de depresión en redes sociales considerando información del perfil de los usuarios**. Tesis de maestría. INAOE. Septiembre 29, 2020.

What about Mexican users?

- Depression detection in Spanish language
  - There *are not labeled data* for Mexican users
- Two work directions:
  - To build our own datasets
  - To apply crosslingual and transfer learning techniques to use existing English resources



Learning bilingual word embeddings with (almost) no bilingual data. Artetxe, M., Labaka, G., and Agirre, E. **ACL 2017**.

Thank you for your attention!

Work in collaboration with:

Mario Ezra Aragón (**INAOE**)
Luis Carlos González (**U. A. Chihuahua**)
A. Pastor López-Monroy (**CIMAT**)

Manuel Montes y Gómez
mmontesg@inaoep.mx
http://ccc.inaoep.mx/~mmontesg