Universität Stuttgart
Institut für
Maschinelle Sprachverarbeitung

Computational Natural Language Understanding: Use cases in the life sciences and psychology

Inaugural Lecture

November 13, 2020

Roman Klinger
roman.klinger@ims.uni-stuttgart.de

@roman_klinger romanklinger
http://www.romanklinger.de/

Universität Stuttgart
Institut für
Maschinelle Sprachverarbeitung

Computational Natural Language Understanding: Use cases in the life sciences and psychology

Inaugural Lecture

November 13, 2020

Roman Klinger
roman.klinger@ims.uni-stuttgart.de

@roman_klinger romanklinger
http://www.romanklinger.de/

Purpose of this talk:
(1) Introduce myself to colleagues*
(2) Identify potential collaborations

* Students/Faculty, including people outside IMS

About Myself (and Stuttgart)

About Myself (and Stuttgart)

- 1999–2006: Studies at University of Dortmund:
  Computer science with minor psychology

UNIVERSITÄT DORTMUND
FACHBEREICH INFORMATIK

Roman Klinger

Komposition von Musik mit Methoden der Computational Intelligence

– Diplomarbeit –

1. Juni 2006

Lehrstuhl 11
Computational Intelligence
Fachbereich Informatik
Universität Dortmund

Gutachter:
Prof. Dr. G. Rudolph
Dr. L. Hildebrand

About Myself (and Stuttgart)

- 1999–2006: Studies at University of Dortmund:
  Computer science with minor psychology
- 2006–2010: Doctoral studies at Fraunhofer SCAI, St. Augustin:
  Biomedical text mining, machine learning

Conditional Random Fields for
Named Entity Recognition

Feature Selection and Optimization in
Biology and Chemistry

Dissertation

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

der Technischen Universität Dortmund
an der Fakultät für Informatik
von

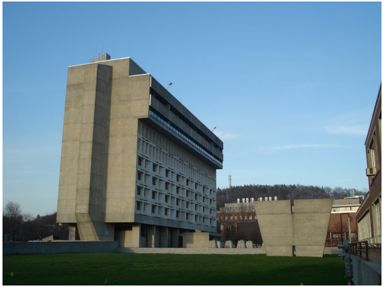
Roman Klinger

Dortmund
2011

Once had a project meeting in a hotel in Stuttgart.

About Myself (and Stuttgart)

- 1999–2006: Studies at University of Dortmund:
  Computer science with minor psychology
- 2006–2010: Doctoral studies at Fraunhofer SCAI, St. Augustin:
  Biomedical text mining, machine learning
- 2010, 2013: Research visits at UMass Amherst:
  Probabilistic machine learning, MCMC inference



Campus Center at UMass, designed by Marcel Breuer, who also designed furniture in the Weissenhof Estate in Stuttgart.

About Myself (and Stuttgart)

- 1999–2006: Studies at University of Dortmund:
  Computer science with minor psychology
- 2006–2010: Doctoral studies at Fraunhofer SCAI, St. Augustin:
  Biomedical text mining, machine learning
- 2010, 2013: Research visits at UMass Amherst:
  Probabilistic machine learning, MCMC inference
- 2011–2012: Postdoc at Fraunhofer SCAI:
  Social media mining, eGovernment

The group at Fraunhofer was connected to Uni Bonn through an institute which was founded through the Berlin-Bonn Act. If Stuttgart won in 1948 to be the capital, that might not have existed.

About Myself (and Stuttgart)

- 1999–2006: Studies at University of Dortmund:
  Computer science with minor psychology
- 2006–2010: Doctoral studies at Fraunhofer SCAI, St. Augustin:
  Biomedical text mining, machine learning
- 2010, 2013: Research visits at UMass Amherst:
  Probabilistic machine learning, MCMC inference
- 2011–2012: Postdoc at Fraunhofer SCAI:
  Social media mining, eGovernment
- 2013–2014: Postdoc at Bielefeld University:
  Sentiment analysis, opinion mining

My postdoc adviser Philipp Cimiano
studied in Stuttgart.

About Myself (and Stuttgart)

- 1999–2006: Studies at University of Dortmund:
  Computer science with minor psychology
- 2006–2010: Doctoral studies at Fraunhofer SCAI, St. Augustin:
  Biomedical text mining, machine learning
- 2010, 2013: Research visits at UMass Amherst:
  Probabilistic machine learning, MCMC inference
- 2011–2012: Postdoc at Fraunhofer SCAI:
  Social media mining, eGovernment
- 2013–2014: Postdoc at Bielefeld University:
  Sentiment analysis, opinion mining
- 2015: Co-Founder of Semalytix GmbH (exit 2020)

Cofounder Matthias Hartung lived in Stuttgart.

About Myself (and Stuttgart)

- 1999–2006: Studies at University of Dortmund:
  Computer science with minor psychology
- 2006–2010: Doctoral studies at Fraunhofer SCAI, St. Augustin:
  Biomedical text mining, machine learning
- 2010, 2013: Research visits at UMass Amherst:
  Probabilistic machine learning, MCMC inference
- 2011–2012: Postdoc at Fraunhofer SCAI:
  Social media mining, eGovernment
- 2013–2014: Postdoc at Bielefeld University:
  Sentiment analysis, opinion mining
- 2015: Co-Founder of Semalytix GmbH (exit 2020)
- 2014–2015: Visiting professor at Uni Stuttgart
- 2015–: (Senior) Lecturer at IMS
- 2020: Habilitation in Computer Science:
  Structured Modelling of Affect in Text

About Myself (and Stuttgart)

- 1999–2006: Studies at University of Dortmund:
  Computer science with minor psychology
- 2006–2010: Doctoral studies at Fraunhofer SCAI, St. Augustin:
  Biomedical text mining, machine learning
- 2010, 2013: Research visits at UMass Amherst:
  Probabilistic machine learning, MCMC inference
- 2011–2012: Postdoc at Fraunhofer SCAI:
  Social media mining, eGovernment
- 2013–2014: Postdoc at Bielefeld University:
  Sentiment analysis, opinion mining
- 2015: Co-Founder of Semalytix GmbH (exit 2020)
- 2014–2015: Visiting professor at Uni Stuttgart
- 2015–: (Senior) Lecturer at IMS
- 2020: Habilitation in Computer Science:
  Structured Modelling of Affect in Text

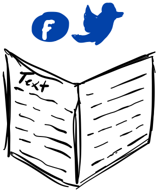
Outline

1. Introduction
2. Biomedical Text Understanding
3. Text Understanding Regarding Psychological Concepts: Emotions
4. Conclusion & Vision

Outline

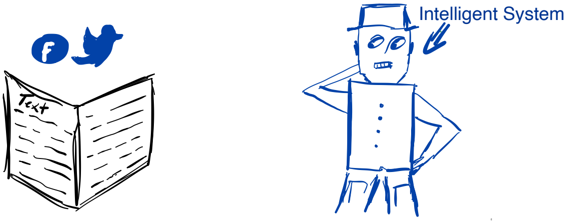
1. Introduction
2. Biomedical Text Understanding
3. Text Understanding Regarding Psychological Concepts: Emotions
4. Conclusion & Vision

Language Understanding

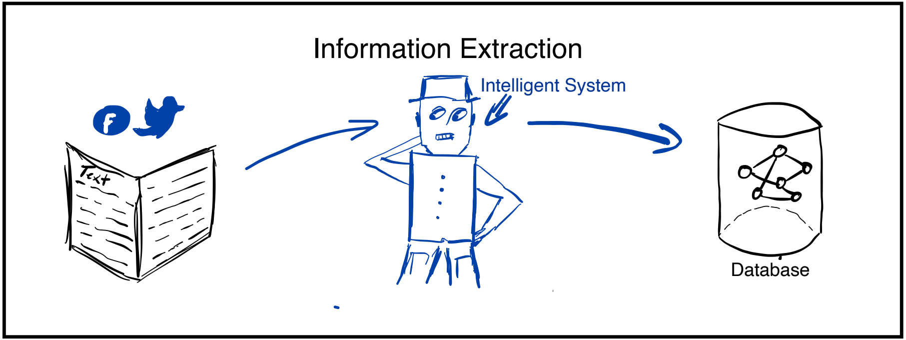


Language Understanding

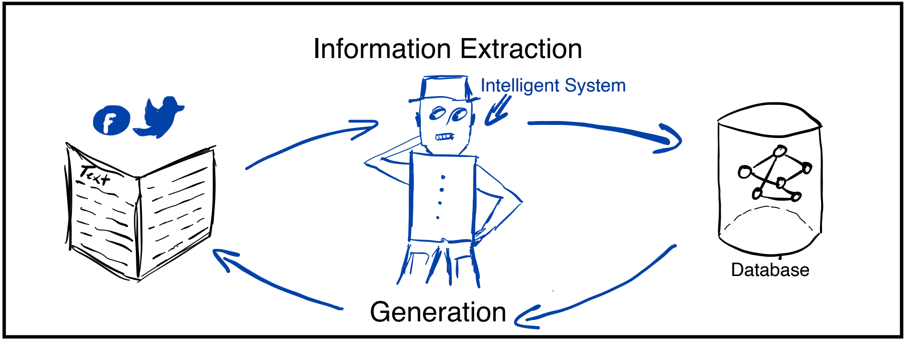
Information Extraction



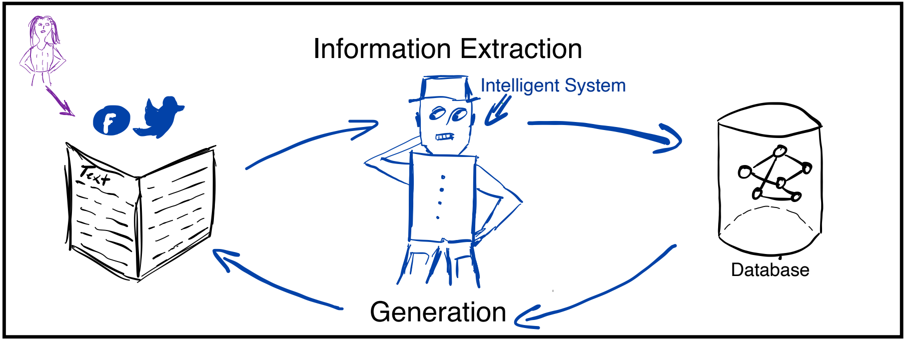
Language Understanding



Language Understanding

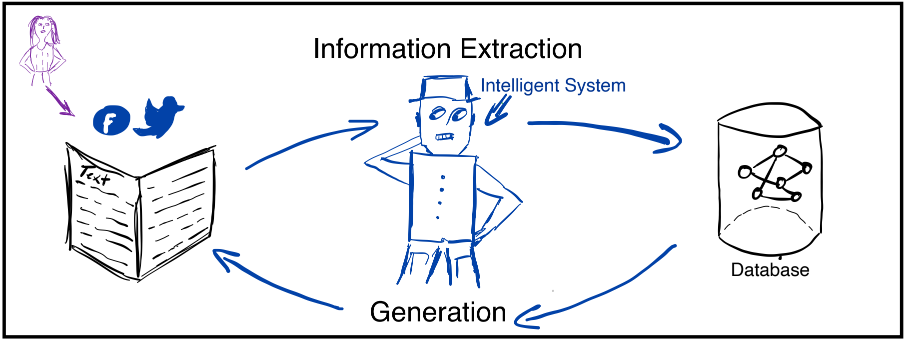


Language Understanding



- Challenges:
  1. Interpret and structure propositional knowledge/statements
  2. Infer properties about author of message

Language Understanding



- Challenges:
  1. Interpret and structure propositional knowledge/statements
  2. Infer properties about author of message
- Two case-studies: Biomedical Information Extraction and Emotion Analysis

Goal of this lecture

Goal of this lecture

- Outline approaches to information extraction, highlight particularities of biomedical NLP and emotion analysis

Goal of this lecture

- Outline approaches to information extraction, highlight particularities of biomedical NLP and emotion analysis
- Highlight differences between text genres/domains and particular challenges

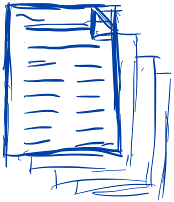
Goal of this lecture

- Outline approaches to information extraction, highlight particularities of biomedical NLP and emotion analysis
- Highlight differences between text genres/domains and particular challenges
- Discuss methodological implications for extraction tasks of different types

Goal of this lecture

- Outline approaches to information extraction, highlight particularities of biomedical NLP and emotion analysis
- Highlight differences between text genres/domains and particular challenges
- Discuss methodological implications for extraction tasks of different types
- I'll let you know at the end how these clearly very different topics can come together in applications.

Concept Identification



$\rightarrow$

Concept$_1$

Concept$_2$

Concept$_3$

Concept$_4$

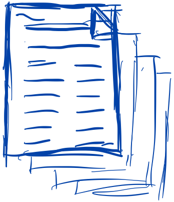
Concept$_5$

Concept$_6$

$\vdots$

Concept$_n$

Concept Identification



Concept_1

Concept_2

Concept_3

$\boxed{\text{Concept}_4}$

Concept_5

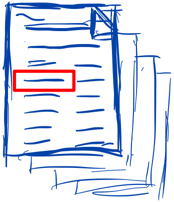
Concept_6

$\vdots$

Concept_n

- Identify concept: sufficient for retrieval

Concept Identification



- Identify concept: sufficient for retrieval
- Identify mention position:
  nice to have for further analysis tasks

Concept_1

Concept_2

Concept_3

$\boxed{\text{Concept}_4}$

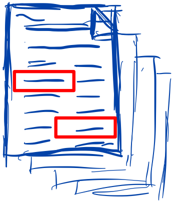
Concept_5

Concept_6

$\vdots$

Concept_n

Concept Identification



Concept_1

Concept_2

Concept_3

$\boxed{\text{Concept}_4}$

Concept_5

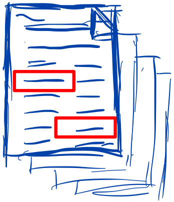
$\boxed{\text{Concept}_6$}

$\vdots$

Concept_n

- Identify concept: sufficient for retrieval
- Identify mention position:
  nice to have for further analysis tasks
- Multiple concepts can be associated with one document

Concept Identification



Concept$_1$
Concept$_2$
Concept$_3$
Concept$_4$
Concept$_5$
Concept$_6$
⋮
Concept$_n$

- Identify concept: sufficient for retrieval
- Identify mention position:
  nice to have for further analysis tasks
- Multiple concepts can be associated with one document
- (I am mixing NER, Entity Linking, and Document Classification here.)

Outline

BioNLP and Medical NLP

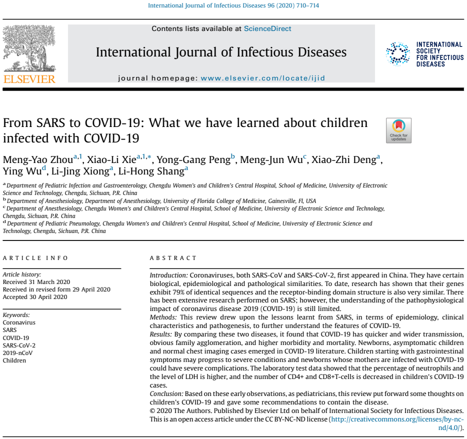
- Automatically extract information from texts in the life science domain

BioNLP and Medical NLP

- Automatically extract information from texts in the life science domain
- A lot of information is hidden in text.

BioNLP and Medical NLP

- Automatically extract information from texts in the life science domain
- A lot of information is hidden in text.
  - Scientific papers (from PubMed)


International Journal of Infectious Diseases 96 (2020) 710–714

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

ELSEVIER

INTERNATIONAL SOCIETY FOR INFECTIOUS DISEASES

From SARS to COVID-19: What we have learned about children infected with COVID-19

Meng-Yao Zhou[a,1], Xiao-Li Xie[a,1,*], Yong-Gang Peng[b], Meng-Jun Wu[c], Xiao-Zhi Deng[a], Ying Wu[d], Li-Jing Xiong[a], Li-Hong Shang[a]

[a] Department of Pediatric Infection and Gastroenterology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China
[b] Department of Anesthesiology, Department of Anesthesiology, University of Florida College of Medicine, Gainesville, Fl, USA
[c] Department of Anesthesiology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China
[d] Department of Pediatric Pneumology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China

ARTICLE INFO

Article history:
Received 31 March 2020
Received in revised form 29 April 2020
Accepted 30 April 2020

Keywords:
Coronavirus
SARS
COVID-19
SARS-CoV-2
2019-nCoV
Children

ABSTRACT

Introduction: Coronaviruses, both SARS-CoV and SARS-CoV-2, first appeared in China. They have certain biological, epidemiological and pathological similarities. To date, research has shown that their genes exhibit 79% of identical sequences and the receptor-binding domain structure is also very similar. There has been extensive research performed on SARS; however, the understanding of the pathophysiological impact of coronavirus disease 2019 (COVID-19) is still limited.
Methods: This review drew upon the lessons learnt from SARS, in terms of epidemiology, clinical characteristics and pathogenesis, to further understand the features of COVID-19.
Results: By comparing these two diseases, it found that COVID-19 has quicker and wider transmission, obvious family agglomeration, and higher morbidity and mortality. Newborns, asymptomatic children and normal chest imaging cases emerged in COVID-19 literature. Children starting with gastrointestinal symptoms may progress to severe conditions and newborns whose mothers are infected with COVID-19 could have severe complications. The laboratory test data showed that the percentage of neutrophils and the level of LDH is higher, and the number of CD4+ and CD8+T-cells is decreased in children's COVID-19 cases.
Conclusion: Based on these early observations, as pediatricians, this review put forward some thoughts on children's COVID-19 and gave some recommendations to contain the disease.
© 2020 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

BioNLP and Medical NLP

- Automatically extract information from texts in the life science domain
- A lot of information is hidden in text.
  - Scientific papers (from PubMed)
  - Discharge letters

Letter Date: *xx/xx/xxxx*
Reference:
Dictated Date: *xx/xx/xxxx*
Transcribed Date: *xx/xx/xxxx*
PATIENT; D.O.B: ..; CHI: .

Admission: *Specialty - .; Ward – xx*
Consultant: *Date of Admission - xx/xx/xxxx*
Date of Discharge *- xx/xx/xxxx;* ***Discharged to:*** *[. .]***Follow Up:*** *[ ]*

Clinical Comments:
Diagnosis: Musculoskeletal chest pain
Ischaemic heart disease
Type II diabetes mellitus
Hypertension
Previous CVA
Obesity
This [..] year old woman was admitted with a complaint of recurrent [chest pain]. There is a background of ischaemic heart disease with previous [. .] myocardial infarction and [. . .]
Other history is of hypertension, cerebral vascular disease, type II diabetes mellitus and obesity. Cardiac examination [. .] ECG showed sinus rhythm with old [. .] infarction. There were no sequential changes and troponin was not raised.
I felt that her symptoms were consistent with musculoskeletal origin. [. .].

Yours sincerely,

Dr [. .]

https://www.isdscotland.org/Products-and-Services/Terminology-Services/Information-for-Clinicians/docs/Discharge-summary-examples_final_CF02.pdf

BioNLP and Medical NLP

- Automatically extract information from texts in the life science domain
- A lot of information is hidden in text.
  - Scientific papers (from PubMed)
  - Discharge letters
  - Documentations of clinical trials

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: https://www.coronavirus.gov.
Get the latest research information from NIH: https://www.nih.gov/coronavirus.

NIH U.S. National Library of Medicine
ClinicalTrials.gov

Trial record **10 of 3541** for: covid

Previous Study | Return to List | Next Study

Effect of COVID-19 Pandemic on Pediatric Cancer Care

The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Know the risks and potential benefits of clinical studies and talk to your health care provider before participating. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT04374838

https:
//clinicaltrials.gov/ct2/show/NCT04374838?cond=covid&draw=2&rank=10

BioNLP and Medical NLP

- Automatically extract information from texts in the life science domain
- A lot of information is hidden in text.
  - Scientific papers (from PubMed)
  - Discharge letters
  - Documentations of clinical trials
- Entity classes of interest:

BioNLP and Medical NLP

- Automatically extract information from texts in the life science domain
- A lot of information is hidden in text.
  - Scientific papers (from PubMed)
  - Discharge letters
  - Documentations of clinical trials
- Entity classes of interest:
  - Gene/names, mutations, species

- BRCA2
- Tinman
- Swiss cheese
- INDY
- superman
- Sonic hedgehog
- Barbie and Ken
- Dracula
- Cheap date
- Dreadlocks

https://thenode.biologists.com/whats-your-favourite-gene/discussion/

BioNLP and Medical NLP

- Automatically extract information from texts in the life science domain
- A lot of information is hidden in text.
  - Scientific papers (from PubMed)
  - Discharge letters
  - Documentations of clinical trials
- Entity classes of interest:
  - Gene/names, mutations, species
  - Chemical compounds, drugs, treatments

- Aspirin
- ACC
- Polopiryna
- 2-acetoxybenzoic acid
- O=C(C)Oc1ccccc1C(=O)O
- C9H8O4

- radiotherapy
- psychotherapy
- physical therapy
- homoeopathy

BioNLP and Medical NLP

- Automatically extract information from texts in the life science domain
- A lot of information is hidden in text.
  - Scientific papers (from PubMed)
  - Discharge letters
  - Documentations of clinical trials
- Entity classes of interest:
  - Gene/names, mutations, species
  - Chemical compounds, drugs, treatments
  - Diseases, medical conditions, adverse effects

- COVID19, corona
- cancer, neoplasm,
- flu, cold, influenza
- patella fracture
- headache, insomnia, vomiting
- death

BioNLP and Medical NLP

- Automatically extract information from texts in the life science domain
- A lot of information is hidden in text.
    - Scientific papers (from PubMed)
    - Discharge letters
    - Documentations of clinical trials
- Entity classes of interest:
    - Gene/names, mutations, species
    - Chemical compounds, drugs, treatments
    - Diseases, medical conditions, adverse effects
    - ...

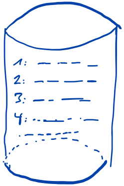
Example: How to find drug names and chemical compounds?

Idea:

Example: How to find drug names and chemical compounds?

Idea:

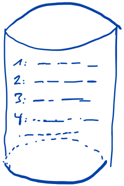
- Let's check for existing data bases

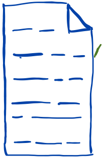


Example: How to find drug names and chemical compounds?

Idea:

- Let's check for existing data bases
- Implement a (fuzzy) dictionary-matching algorithm



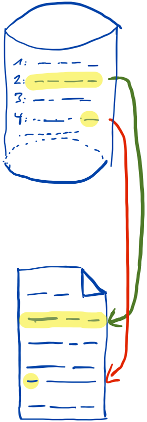


Example: How to find drug names and chemical compounds?

Idea:

- Let's check for existing data bases
- Implement a (fuzzy) dictionary-matching algorithm

⇒ Find mentions and link to databases in one step



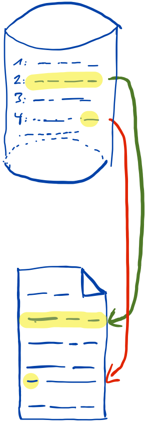
Example: How to find drug names and chemical compounds?

Idea:

- Let's check for existing data bases
- Implement a (fuzzy) dictionary-matching algorithm

⇒ Find mentions and link to databases in one step

⇒ Directly find chemical compound



Example: How to find drug names and chemical compounds?

Idea:

- Let's check for existing data bases
- Implement a (fuzzy) dictionary-matching algorithm

⇒ Find mentions and link to databases in one step

⇒ Directly find chemical compound

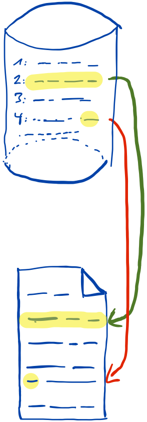
Chemical Names: Terminological Resources and Corpora Annotation

Corinna Kolářik*†, Roman Klinger†,
Christoph M. Friedrich†, Martin Hofmann-Apitius*†, and Juliane Fluck†

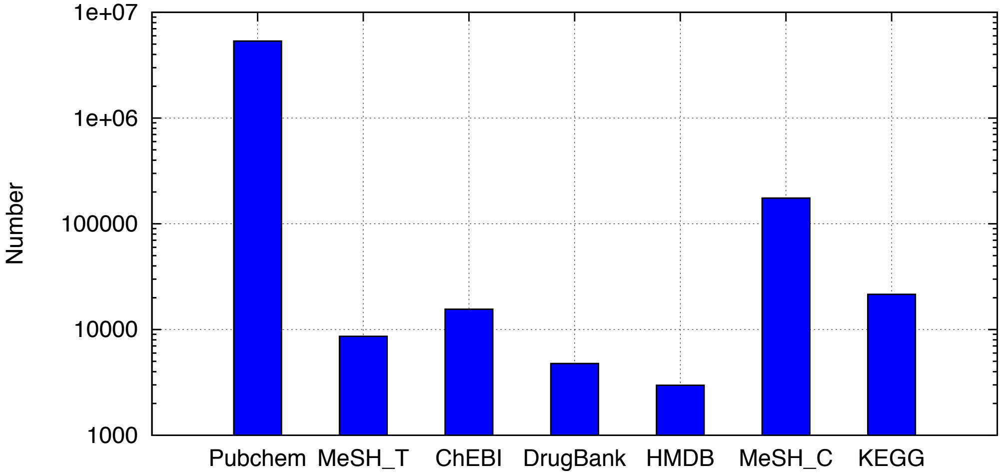
†Fraunhofer Institute Algorithms
and Scientific Computing (SCAI)
Department of Bioinformatics
Schloß Birlinghoven
53574 Sankt Augustin, Germany

*Bonn-Aachen International Center
for Information Technology (B-IT)
Department of Applied Life Science Informatics
Dahlmannstrasse 2
D-53113 Bonn, Germany

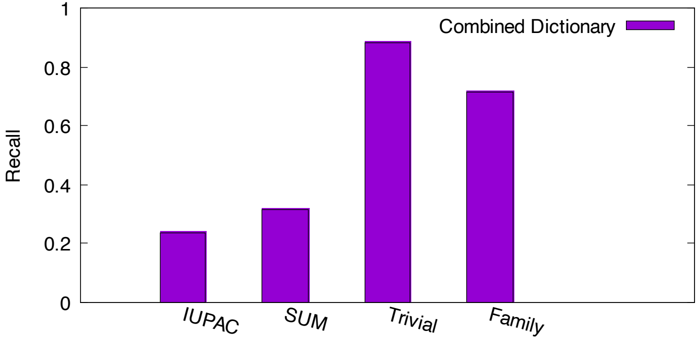
corinna.kolarik@scai.fhg.de, roman.klinger@scai.fhg.de,
christoph.friedrich@scai.fhg.de, martin.hofmann-apitius@scai.fhg.de, juliane.fluck@scai.fhg.de



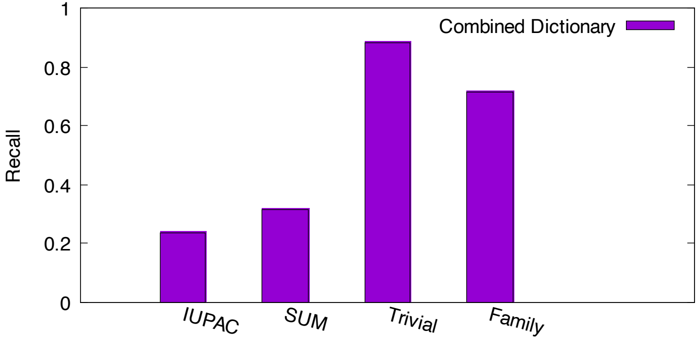
Chemical compound databases: Size



Chemical compound databases: Recall on annotated corpus



Chemical compound databases: Recall on annotated corpus



- But: 0.13 Precision.

Example results

Example results

DEPARTMENTS: NURSING AND THE ARTS

Empowering Women Since 1912: The Girl Scouts of America

YOUNG-MASON, JEANINE EdD, RN, CS, FAAN

Section Editor(s): Young-Mason, Jeanine EdD, RN, CS, FAAN **Author Information**

Clinical Nurse Specialist: July/August 2012 - Volume 26 - Issue 4 - p 227-228
doi: 10.1097/NUR.0b013e31825aea30

Example results

DEPARTMENTS: NURSING AND THE ARTS

Empowering Women Since 1912: The Girl Scouts of America

YOUNG-MASON, JEANINE EdD, RN, CS, FAAN

Section Editor(s): Young-Mason, Jeanine EdD, RN, CS, FAAN **Author Information**

Clinical Nurse Specialist: July/August 2012 - Volume 26 - Issue 4 - p 227-228
doi: 10.1097/NUR.0b013e31825aea30

Clin Exp Immunol. 2019 Apr; 196(1): 28–38. PMCID: PMC6422647
Published online 2019 Feb 27. doi: 10.1111/cei.13265 PMID: 30697704

Exploring immunomodulation by endocrine changes in Lady Windermere syndrome

M. R. Holt,[1,2] J. J. Miles,[3] W. J. Inder,[1,4] and R. M. Thomson[1,2]

Example results

DEPARTMENTS: NURSING AND THE ARTS

Empowering Women Since 1912: The Girl Scouts of America

YOUNG-MASON, JEANINE EdD, RN, CS, FAAN

Section Editor(s): Young-Mason, Jeanine EdD, RN, CS, FAAN **Author Information**

Clinical Nurse Specialist: July/August 2012 - Volume 26 - Issue 4 - p 227-228
doi: 10.1097/NUR.0b013e31825aea30

Clin Exp Immunol. 2019 Apr; 196(1): 28–38. PMCID: PMC6422647
Published online 2019 Feb 27. doi: 10.1111/cei.13265 PMID: 30697704

Exploring immunomodulation by endocrine changes in Lady Windermere syndrome

M. R. Holt,[1,2] J. J. Miles,[3] W. J. Inder,[1,4] and R. M. Thomson[1,2]

Review > Res Microbiol. 2015 Dec;166(10):782-95. doi: 10.1016/j.resmic.2015.09.002. Epub 2015 Sep 25.

Snow and ice ecosystems: not so extreme

Lorrie Maccario[1], Laura Sanguino[1], Timothy M Vogel[1], Catherine Larose[2]

Affiliations + expand

PMID: 26408452 DOI: 10.1016/j.resmic.2015.09.002

Example results

DEPARTMENTS: NURSING AND THE ARTS

Empowering Women Since 1912: The Girl Scouts of America

YOUNG-MASON, JEANINE EdD, RN, CS, FAAN

Section Editor(s): Young-Mason, Jeanine EdD, RN, CS, FAAN **Author Information**

Clinical Nurse Specialist: July/August 2012 - Volume 26 - Issue 4 - p 227-228
doi: 10.1097/NUR.0b013e31825aea30

Clin Exp Immunol. 2019 Apr; 196(1): 28–38. PMCID: PMC6422647
Published online 2019 Feb 27. doi: 10.1111/cei.13265 PMID: 30697704

Exploring immunomodulation by endocrine changes in Lady Windermere syndrome

M. R. Holt,[1,2] J. J. Miles,[3] W. J. Inder,[1,4] and R. M. Thomson[1,2]

Review > Res Microbiol. 2015 Dec;166(10):782-95. doi: 10.1016/j.resmic.2015.09.002. Epub 2015 Sep 25.

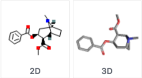
Snow and ice ecosystems: not so extreme

Lorrie Maccario[1], Laura Sanguino[1], Timothy M Vogel[1], Catherine Larose[2]

Affiliations + expand

PMID: 26408452 DOI: 10.1016/j.resmic.2015.09.002

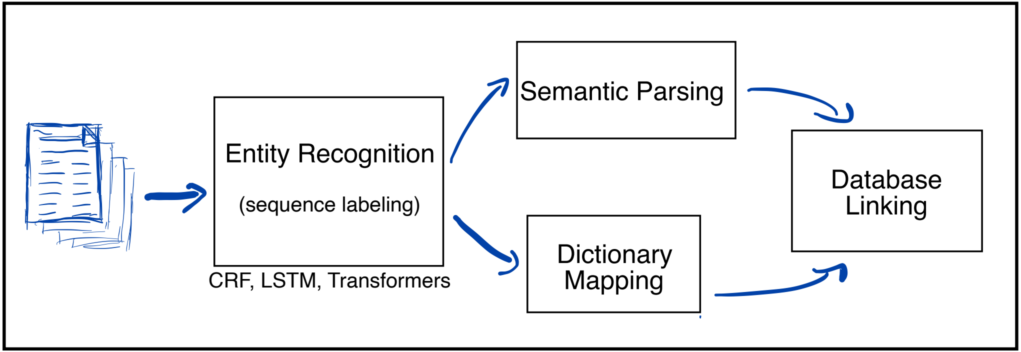
Cocaine

PubChem CID:	446220
Structure:	 2D 3D Find Similar Structures
Chemical Safety:	 Acute Toxic Irritant Laboratory Chemical Safety Summary (LCSS) Datasheet
Molecular Formula:	$C_{17}H_{21}NO_4$
Synonyms:	cocaine Kokain Neurocaine Cocain L-Cocaine More...

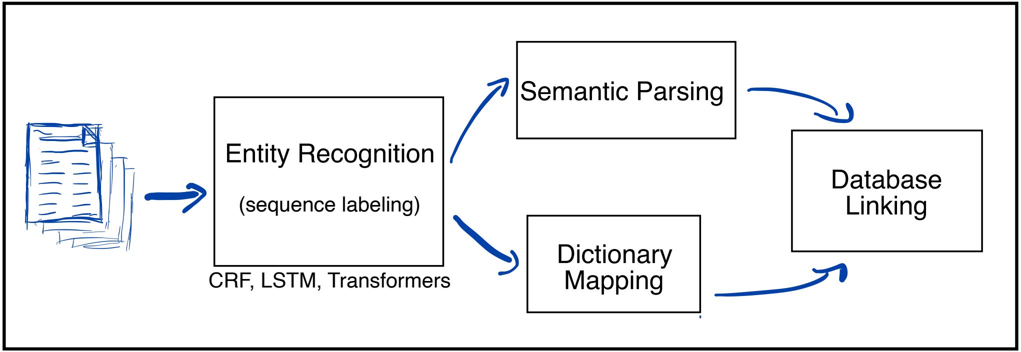
Cocaine Synonyms

cocaine Kokain Neurocaine Cocain L-Cocaine Cocaina beta-Cocain (-)-Cocaine Methyl Benzoylecgonine l-Cocain Benzoylmethylecgonine Leaf Dama blanca Pimp's drug Cocaine free base 1-Cocaine **White girl or lady** cocainum Star-spangled powder Cocaine, l- Eritroxilina Erytroxylin Kokayeen Bernies Burese Corine 50-36-2 Kokan Coke UNII-I5Y540LHVR Bernice Cholly Cecil Flake Blow **Girl** Lady Rock **Snow** Toot Happy trails Green gold Happy dust Nose candy Gold dust Star dust CHEBI:27958 2-beta-Carbomethoxy-3-beta-benzoxytropane HSDB 6469 C"Carrie I5Y540LHVR 2-beta-Tropanecarboxylic acid, 3-beta-hydroxy-, methyl ester, benzoate (ester) methyl (1R,2R,3S,5S)-3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane-2-carboxylate Crack cocaine Methyl 3beta-hydroxy-1alphaH,5alphaH-tropane-2beta-carboxylate benzoate (ester) COC Ecgonine, methyl ester, benzoate (ester) Jam Crack 3-Tropanylbenzoate-2-carboxylic acid methyl ester 2beta-Carbomethoxy-3beta-benzoxytropane 1-alpha-H,5-alpha-H-Tropane-2-beta-carboxylic acid, 3-beta-hydroxy-, methyl ester, benzoate 2-Methyl-3beta-hydroxy-1alphaH,5alphaH-tropane-2beta-carboxylate benzoate (ester) 3-(Benzoyloxy)-8-methyl-8-azabicyclo-(3.2.1)octane-2-carboxylic acid methyl ether 3beta-Hydroxy-1alphaH,5alphaH-tropane-2beta-carboxylic acid methyl ester benzoate methyl (1S,3S,4R,5R)-3-benzoyloxy-8-methyl-8-azabicyclo[3.2.1]octane-4-carboxylate methyl [1R-(exo,exo)]-3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane-2-carboxylate Methyl 3-beta-hydroxy-1-alpha-H,5-alpha-H-tropane-2-beta-carboxylate benzoate (ester) (1R,2R,3S,5S)-2-Methoxycarbonyltropan-3-yl benzoate Blow [Street Name] **Girl Lady** Rock Toot Cecil Flake Sleighride Badrock Bazooka Bernice **Blizzard** Cabello Charlie Cocktail Goofball Moonrocks Blast Candy Caviar Freeze Heaven Snort Trails Coca Cola Hell Toke Yeyo Bouncing Powder Chicken Scratch Happy powder EINECS 200-032-7 Florida Snow Sweet Stuff Gold dust [Street Name] Prime Time C Carrie Happy dust [Street Name] 8-Azabicyclo(3.2.1)octane-2-carboxylic acid, 3-(benzoyloxy)-8-methyl-, methyl ester, (1R-(exo,exo))- Foo Foo Kibbles n' Bits Snow (birds) G-Rock [1R-(exo,exo)]-3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane-2-carboxylic acid, methyl ester methyl (1R,2R,3S,5S)-8-methyl-3-[(phenylcarbonyl)oxy]-8-azabicyclo[3.2.1]octane-2-carboxylate Cholly [Street Name] Cocaine [USP:BAN] Star dust [Street Name] Green gold [Street Name] DEA No. 9041 Happy trails [Street Name] Line Cocaine (-) 1i7z Epitope ID:158626 SCHEMBL21930 CHEMBL370805 GTPL2286 IDS-NC-004 DTXSID2038443 BDBM22418 (1R,2R,3S,5S)-2-(methoxycarbonyl)tropan-3-yl benzoate 1q72 Cocaine 0.1 mg/ml in Acetonitrile Cocaine 1.0 mg/ml in Acetonitrile ZINC3875336 RX0041 AKOS015965554 DB00907 RX-0041 C01416 Q41576 (1R,5S,8R)-2beta-(Methoxycarbonyl)-3beta-(benzoyloxy)tropane cocaine hydrochloride;Cocaine hydrochloride;(-)-Cocaine hydrochloride [1R-(exo,exo)]-3-(Benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane-2-carboxylic Acid Cocaine solution, 1.0 mg/mL in acetonitrile, ampule of 1 mL, certified reference material methyl (2R,3S)-3-(benzoyloxy)-8-methyl-8-azabicyclo[3.2.1]octane-2-carboxylate (1beta,5beta,8-anti)-3beta-Benzoyloxy-8-methyl-8-azabicyclo[3.2.1]octane-2beta-carboxylic acid methyl ester 1-alpha-H,5-alpha-H-Tropane-2-beta-carboxylic acid, 3-beta-hydroxy-, methyl ester, benzoate (ester) (8CI) 8-Azabicyclo[3.2.1]octane-2-carboxylic acid, 3-(benzoyloxy)-8-methyl-, methyl ester, (1R,2R,3S,5S)-(9CI)

Chemical NER Pipeline

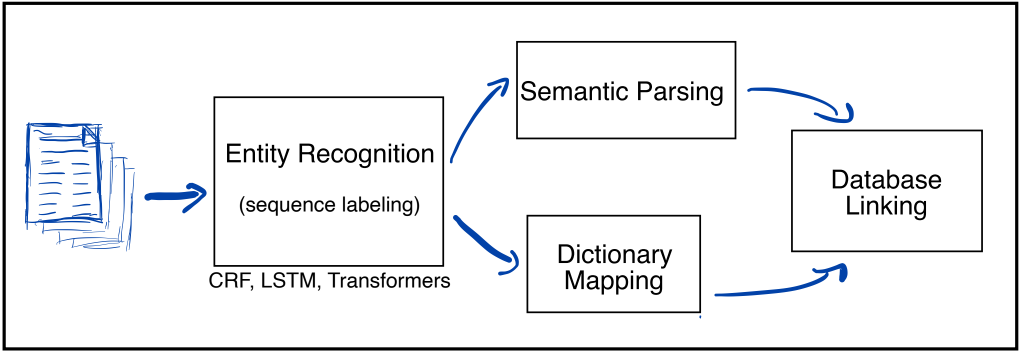


Chemical NER Pipeline



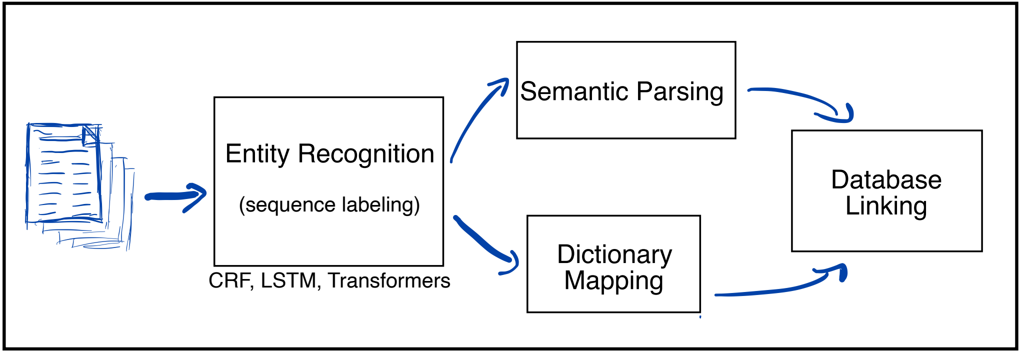
- State-of-the-art: .92 F_1 for entity recognition (BioBERT, Lee et al. 2020)

Chemical NER Pipeline



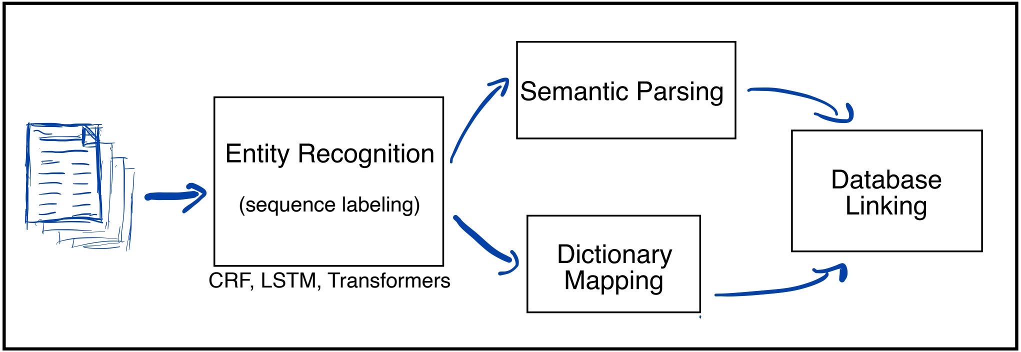
- State-of-the-art: .92 F_1 for entity recognition (BioBERT, Lee et al. 2020)
- Open research topics:

Chemical NER Pipeline



- State-of-the-art: .92 F_1 for entity recognition (BioBERT, Lee et al. 2020)
- Open research topics:
  - Joint models for linking and NER

Chemical NER Pipeline



- State-of-the-art: .92 F_1 for entity recognition (BioBERT, Lee et al. 2020)
- Open research topics:
  - Joint models for linking and NER
  - Multi-task learning for relations and entities

A glimpse on disease names

A glimpse on disease names

- Results are similar for disease names:

A glimpse on disease names

- Results are similar for disease names:
    - Several combined dictionaries on manually annotated paper abstracts: .19 Precision, .76 Recall

An Empirical Evaluation of Resources for the Identification of Diseases and Adverse Effects in Biomedical Literature

Harsha Gurulingappa*†, Roman Klinger*, Martin Hofmann-Apitius*†, and Juliane Fluck*

*Fraunhofer Institute for Algorithms and Scientific Computing
Schloss Birlinghoven, 53754 Sankt Augustin, Germany
†Bonn-Aachen International Center for Information Technology
Dahlmannstraße 2, 53113 Bonn, Germany
harsha.gurulingappa@scai-extern.fraunhofer.de,
{roman.klinger, martin.hofmann-apitius, and juliane.fluck}@scai.fraunhofer.de

A glimpse on disease names

- Results are similar for disease names:
  - Several combined dictionaries on manually annotated paper abstracts: .19 Precision, .76 Recall
  - Joint recognition and normalization works pretty well though (.8 F_1)

BIOINFORMATICS ORIGINAL PAPER Vol. 29 no. 22 2013, pages 2909–2917
doi:10.1093/bioinformatics/btt474

Data and text mining Advance Access publication August 21, 2013

DNorm: disease name normalization with pairwise learning to rank

Robert Leaman[1,2], Rezarta Islamaj Doğan[1] and Zhiyong Lu[1,*]
[1]National Center for Biotechnology Information, 8600 Rockville Pike, Bethesda, MD 20894, USA and [2]Department of Biomedical Informatics, Arizona State University, 13212 East Shea Blvd, Scottsdale, AZ 85259, USA
Associate Editor: Jonathan Wren

A glimpse on disease names

- Results are similar for disease names:
  - Several combined dictionaries on manually annotated paper abstracts: .19 Precision, .76 Recall
  - Joint recognition and normalization works pretty well though (.8 F_1)
- Disease name recognition is challenging when compared between social media and scientific text

On the Semantic Similarity of Disease Mentions in MEDLINE® and Twitter

Camilo Thorne and Roman Klinger

Institut für Maschinelle Sprachverarbeitung (IMS), University of Stuttgart
{camilo.thorne,roman.klinger}@ims.uni-stuttgart.de

A glimpse on disease names

- Results are similar for disease names:
  - Several combined dictionaries on manually annotated paper abstracts: .19 Precision, .76 Recall
  - Joint recognition and normalization works pretty well though (.8 F_1)
- Disease name recognition is challenging when compared between social media and scientific text
  - Frequent names are similarly used, infrequent ones are dissimilar

Table 3: 7 most similar MeSH concepts.

MeSH ID	Similarity	Canonical name
D006526	0.496	Hepatitis C
D005910	0.463	Glioma
D003920	0.459	Diabetes Mellitus
D006521	0.453	Chronic Hepatitis
D000860	0.451	Hypoxia
D003327	0.446	Coronary Disease
D015658	0.445	HIV Infections
...	...	...

Table 4: 6 least similar MeSH concepts.

MeSH ID	Similarity	Canonical name
...	...	...
D015458	0.170	T Cell Leukemia
D002547	0.155	Cerebral Palsy
C536528	0.122	Van der Woude syndrome
C535984	0.116	Congenital bilateral aplasia of vas deferens
D029461	0.109	Sialic Acid Storage Disease
C537666	0.109	BMD

A glimpse on disease names

- Results are similar for disease names:
  - Several combined dictionaries on manually annotated paper abstracts: .19 Precision, .76 Recall
  - Joint recognition and normalization works pretty well though (.8 F_1)
- Disease name recognition is challenging when compared between social media and scientific text
  - Frequent names are similarly used, infrequent ones are dissimilar
  - Main reason: Ambiguous synonyms in dictionary entry

Table 3: 7 most similar MeSH concepts.

MeSH ID	Similarity	Canonical name
D006526	0.496	Hepatitis C
D005910	0.463	Glioma
D003920	0.459	Diabetes Mellitus
D006521	0.453	Chronic Hepatitis
D000860	0.451	Hypoxia
D003327	0.446	Coronary Disease
D015658	0.445	HIV Infections
...	...	...

Table 4: 6 least similar MeSH concepts.

MeSH ID	Similarity	Canonical name
...	...	...
D015458	0.170	T Cell Leukemia
D002547	0.155	Cerebral Palsy
C536528	0.122	Van der Woude syndrome
C535984	0.116	Congenital bilateral aplasia of vas deferens
D029461	0.109	Sialic Acid Storage Disease
C537666	0.109	BMD

A glimpse on disease names

- Results are similar for disease names:
  - Several combined dictionaries on manually annotated paper abstracts: .19 Precision, .76 Recall
  - Joint recognition and normalization works pretty well though (.8 F_1)
- Disease name recognition is challenging when compared between social media and scientific text
  - Frequent names are similarly used, infrequent ones are dissimilar
  - Main reason: Ambiguous synonyms in dictionary entry
  - Research direction: learn to align expert and layperson language

Table 3: 7 most similar MeSH concepts.

MeSH ID	Similarity	Canonical name
D006526	0.496	Hepatitis C
D005910	0.463	Glioma
D003920	0.459	Diabetes Mellitus
D006521	0.453	Chronic Hepatitis
D000860	0.451	Hypoxia
D003327	0.446	Coronary Disease
D015658	0.445	HIV Infections
...	...	...

Table 4: 6 least similar MeSH concepts.

MeSH ID	Similarity	Canonical name
...	...	...
D015458	0.170	T Cell Leukemia
D002547	0.155	Cerebral Palsy
C536528	0.122	Van der Woude syndrome
C535984	0.116	Congenital bilateral aplasia of vas deferens
D029461	0.109	Sialic Acid Storage Disease
C537666	0.109	BMD

Summary

- Biomedical information extraction is commonly formulated as
  1. machine-learning based NER
  2. dictionary-based Named Entity Normalization
  3. (relation extraction)

Summary

- Biomedical information extraction is commonly formulated as
  1. machine-learning based NER
  2. dictionary-based Named Entity Normalization
  3. (relation extraction)
- Biomedical domain is characterized by rich and high-quality resources (mostly)

Summary

- Biomedical information extraction is commonly formulated as
  1. machine-learning based NER
  2. dictionary-based Named Entity Normalization
  3. (relation extraction)
- Biomedical domain is characterized by rich and high-quality resources (mostly)
- Many shared tasks exist for many different entity types:
  - CHEMDNER (Biocreative)
  - BioNLP Infectious Diseases (BioNLP-ST)
  - Drug Adverse Reactions (AMIA 2017)
  - ...

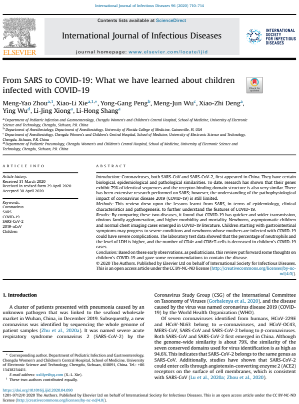
Summary

- Biomedical information extraction is commonly formulated as
  1. machine-learning based NER
  2. dictionary-based Named Entity Normalization
  3. (relation extraction)
- Biomedical domain is characterized by rich and high-quality resources (mostly)
- Many shared tasks exist for many different entity types:
  - CHEMDNER (Biocreative)
  - BioNLP Infectious Diseases (BioNLP-ST)
  - Drug Adverse Reactions (AMIA 2017)
  - ...
- Many entities: 94M in Pubchem (2017)

Summary

- Biomedical information extraction is commonly formulated as
  1. machine-learning based NER
  2. dictionary-based Named Entity Normalization
  3. (relation extraction)
- Biomedical domain is characterized by rich and high-quality resources (mostly)
- Many shared tasks exist for many different entity types:
  - CHEMDNER (Biocreative)
  - BioNLP Infectious Diseases (BioNLP-ST)
  - Drug Adverse Reactions (AMIA 2017)
  - ...
- Many entities: 94M in Pubchem (2017)
- Number of realizations of each entity limited: The challenge is to categorize huge amounts of "classes" to text, though each classification problem is comparably straight-forward. ⇒ Reason for formulation of NER+NEN.

Differences between Scientific Papers and Social Media



International Journal of Infectious Diseases 96 (2020) 710–714

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

ELSEVIER

INTERNATIONAL SOCIETY FOR INFECTIOUS DISEASES

From SARS to COVID-19: What we have learned about children infected with COVID-19

Meng-Yao Zhou[a,1], Xiao-Li Xie[a,1,*], Yong-Gang Peng[b], Meng-Jun Wu[c], Xiao-Zhi Deng[a], Ying Wu[d], Li-Jing Xiong[a], Li-Hong Shang[a]

[a] Department of Pediatric Infection and Gastroenterology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China
[b] Department of Anesthesiology, Department of Anesthesiology, University of Florida College of Medicine, Gainesville, Fl, USA
[c] Department of Anesthesiology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China
[d] Department of Pediatric Pneumology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China

ARTICLE INFO

Article history:
Received 31 March 2020
Received in revised form 29 April 2020
Accepted 30 April 2020

Keywords:
Coronavirus
SARS
COVID-19
SARS-CoV-2
2019-nCoV
Children

ABSTRACT

Introduction: Coronaviruses, both SARS-CoV and SARS-CoV-2, first appeared in China. They have certain biological, epidemiological and pathological similarities. To date, research has shown that their genes exhibit 79% of identical sequences and the receptor-binding domain structure is also very similar. There has been extensive research performed on SARS; however, the understanding of the pathophysiological impact of coronavirus disease 2019 (COVID-19) is still limited.
Methods: This review drew upon the lessons learnt from SARS, in terms of epidemiology, clinical characteristics and pathogenesis, to further understand the features of COVID-19.
Results: By comparing these two diseases, it found that COVID-19 has quicker and wider transmission, obvious family agglomeration, and higher morbidity and mortality. Newborns, asymptomatic children and normal chest imaging cases emerged in COVID-19 literature. Children starting with gastrointestinal symptoms may progress to severe conditions and newborns whose mothers are infected with COVID-19 could have severe complications. The laboratory test data showed that the percentage of neutrophils and the level of LDH is higher, and the number of CD4+ and CD8+T-cells is decreased in children's COVID-19 cases.
Conclusion: Based on these early observations, as pediatricians, this review put forward some thoughts on children's COVID-19 and gave some recommendations to contain the disease.
© 2020 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

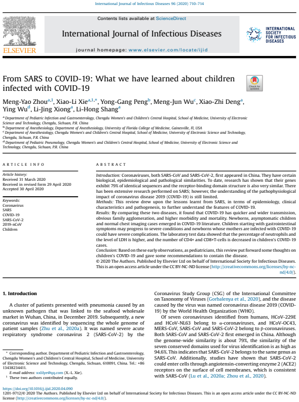
A cluster of patients presented with pneumonia caused by an unknown pathogen that was linked to the seafood wholesale market in Wuhan, China, in December 2019. Subsequently, a new coronavirus was identified by sequencing the whole genome of patient samples (Zhu et al., 2020a). It was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the Coronavirus Study Group (CSG) of the International Committee on Taxonomy of Viruses (Gorbalenya et al., 2020), and the disease caused by the virus was named coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO).

Of seven coronaviruses identified from humans, HCoV-229E and HCoV-NL63 belong to α-coronaviruses, and HCoV-OC43, MERS-CoV, SARS-CoV and SARS-CoV-2 belong to β-coronaviruses. Both SARS-CoV and SARS-CoV-2 first emerged in China. Although the genome-wide similarity is about 79%, the similarity of the seven conserved domains used for virus identification is as high as 94.6%. This indicates that SARS-CoV-2 belongs to the same genus as SARS-CoV. Additionally, studies have shown that SARS-CoV-2 could enter cells through angiotensin-converting enzyme 2 (ACE2) receptors on the surface of cell membranes, which is consistent with SARS-CoV (Lu et al., 2020a; Zhou et al., 2020).

* Corresponding author. Department of Pediatric Infection and Gastroenterology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, 610091, China. Tel.: +86 13438234411.
E-mail address: xxlilye@qq.com (X.-L. Xie).
[1] These two authors contributed equally.

https://doi.org/10.1016/j.ijid.2020.04.090
1201-9712/© 2020 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Differences between Scientific Papers and Social Media



International Journal of Infectious Diseases 96 (2020) 710–714

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

ELSEVIER

INTERNATIONAL SOCIETY FOR INFECTIOUS DISEASES

From SARS to COVID-19: What we have learned about children infected with COVID-19

Meng-Yao Zhou[a,1], Xiao-Li Xie[a,1,*], Yong-Gang Peng[b], Meng-Jun Wu[c], Xiao-Zhi Deng[a], Ying Wu[d], Li-Jing Xiong[a], Li-Hong Shang[a]

[a] *Department of Pediatric Infection and Gastroenterology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China*
[b] *Department of Anesthesiology, Department of Anesthesiology, University of Florida College of Medicine, Gainesville, Fl, USA*
[c] *Department of Anesthesiology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China*
[d] *Department of Pediatric Pneumology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China*

ARTICLE INFO

Article history:
Received 31 March 2020
Received in revised form 29 April 2020
Accepted 30 April 2020

Keywords:
Coronavirus
SARS
COVID-19
SARS-CoV-2
2019-nCoV
Children

ABSTRACT

Introduction: Coronaviruses, both SARS-CoV and SARS-CoV-2, first appeared in China. They have certain biological, epidemiological and pathological similarities. To date, research has shown that their genes exhibit 79% of identical sequences and the receptor-binding domain structure is also very similar. There has been extensive research performed on SARS; however, the understanding of the pathophysiological impact of coronavirus disease 2019 (COVID-19) is still limited.
Methods: This review drew upon the lessons learnt from SARS, in terms of epidemiology, clinical characteristics and pathogenesis, to further understand the features of COVID-19.
Results: By comparing these two diseases, it found that COVID-19 has quicker and wider transmission, obvious family agglomeration, and higher morbidity and mortality. Newborns, asymptomatic children and normal chest imaging cases emerged in COVID-19 literature. Children starting with gastrointestinal symptoms may progress to severe conditions and newborns whose mothers are infected with COVID-19 could have severe complications. The laboratory test data showed that the percentage of neutrophils and the level of LDH is higher, and the number of CD4+ and CD8+T-cells is decreased in children's COVID-19 cases.
Conclusion: Based on these early observations, as pediatricians, this review put forward some thoughts on children's COVID-19 and gave some recommendations to contain the disease.
© 2020 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

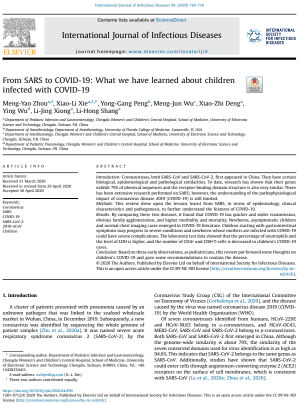
A cluster of patients presented with pneumonia caused by an unknown pathogen that was linked to the seafood wholesale market in Wuhan, China, in December 2019. Subsequently, a new coronavirus was identified by sequencing the whole genome of patient samples (Zhu et al., 2020a). It was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the Coronavirus Study Group (CSG) of the International Committee on Taxonomy of Viruses (Gorbalenya et al., 2020), and the disease caused by the virus was named coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO).

Of seven coronaviruses identified from humans, HCoV-229E and HCoV-NL63 belong to α-coronaviruses, and HCoV-OC43, MERS-CoV, SARS-CoV and SARS-CoV-2 belong to β-coronaviruses. Both SARS-CoV and SARS-CoV-2 first emerged in China. Although the genome-wide similarity is about 79%, the similarity of the seven conserved domains used for virus identification is as high as 94.6%. This indicates that SARS-CoV-2 belongs to the same genus as SARS-CoV. Additionally, studies have shown that SARS-CoV-2 could enter cells through angiotensin-converting enzyme 2 (ACE2) receptors on the surface of cell membranes, which is consistent with SARS-CoV (Lu et al., 2020a; Zhou et al., 2020).

* Corresponding author. Department of Pediatric Infection and Gastroenterology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, 610091, China. Tel.: +86 13438234411.
E-mail address: xxlilye@qq.com (X.-L. Xie).
[1] These two authors contributed equally.

https://doi.org/10.1016/j.ijid.2020.04.090
1201-9712/© 2020 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Differences between Scientific Papers and Social Media



International Journal of Infectious Diseases 96 (2020) 710–714

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

From SARS to COVID-19: What we have learned about children infected with COVID-19

Meng-Yao Zhou[a,1], Xiao-Li Xie[a,1,*], Yong-Gang Peng[b], Meng-Jun Wu[c], Xiao-Zhi Deng[a], Ying Wu[d], Li-Jing Xiong[a], Li-Hong Shang[a]

[a] Department of Pediatric Infection and Gastroenterology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China
[b] Department of Anesthesiology, Department of Anesthesiology, University of Florida College of Medicine, Gainesville, Fl, USA
[c] Department of Anesthesiology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China
[d] Department of Pediatric Pneumology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China

ARTICLE INFO

Article history:
Received 31 March 2020
Received in revised form 29 April 2020
Accepted 30 April 2020

Keywords:
Coronavirus
SARS
COVID-19
SARS-CoV-2
2019-nCoV
Children

ABSTRACT

Introduction: Coronaviruses, both SARS-CoV and SARS-CoV-2, first appeared in China. They have certain biological, epidemiological and pathological similarities. To date, research has shown that their genes exhibit 79% of identical sequences and the receptor-binding domain structure is also very similar. There has been extensive research performed on SARS; however, the understanding of the pathophysiological impact of coronavirus disease 2019 (COVID-19) is still limited.
Methods: This review drew upon the lessons learnt from SARS, in terms of epidemiology, clinical characteristics and pathogenesis, to further understand the features of COVID-19.
Results: By comparing these two diseases, it found that COVID-19 has quicker and wider transmission, obvious family agglomeration, and higher morbidity and mortality. Newborns, asymptomatic children and normal chest imaging cases emerged in COVID-19 literature. Children starting with gastrointestinal symptoms may progress to severe conditions and newborns whose mothers are infected with COVID-19 could have severe complications. The laboratory test data showed that the percentage of neutrophils and the level of LDH is higher, and the number of CD4+ and CD8+T-cells is decreased in children's COVID-19 cases.
Conclusion: Based on these early observations, as pediatricians, this review put forward some thoughts on children's COVID-19 and gave some recommendations to contain the disease.
© 2020 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

A cluster of patients presented with pneumonia caused by an unknown pathogen that was linked to the seafood wholesale market in Wuhan, China, in December 2019. Subsequently, a new coronavirus was identified by sequencing the whole genome of patient samples (Zhu et al., 2020a). It was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the Coronavirus Study Group (CSG) of the International Committee on Taxonomy of Viruses (Gorbalenya et al., 2020), and the disease caused by the virus was named coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO).

Of seven coronaviruses identified from humans, HCoV-229E and HCoV-NL63 belong to α-coronaviruses, and HCoV-OC43, MERS-CoV, SARS-CoV and SARS-CoV-2 belong to β-coronaviruses. Both SARS-CoV and SARS-CoV-2 first emerged in China. Although the genome-wide similarity is about 79%, the similarity of the seven conserved domains used for virus identification is as high as 94.6%. This indicates that SARS-CoV-2 belongs to the same genus as SARS-CoV. Additionally, studies have shown that SARS-CoV-2 could enter cells through angiotensin-converting enzyme 2 (ACE2) receptors on the surface of cell membranes, which is consistent with SARS-CoV (Lu et al., 2020a; Zhou et al., 2020).

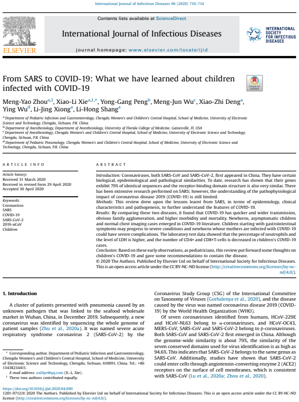
* Corresponding author. Department of Pediatric Infection and Gastroenterology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, 610091, China. Tel.: +86 13438234411.
E-mail address: xxlilye@qq.com (X.-L. Xie).
[1] These two authors contributed equally.

https://doi.org/10.1016/j.ijid.2020.04.090
1201-9712/© 2020 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

⇐

- Trustworthy
- Fact-oriented, precise language
- Slow publication process

Differences between Scientific Papers and Social Media



International Journal of Infectious Diseases 96 (2020) 710–714

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

From SARS to COVID-19: What we have learned about children infected with COVID-19

Meng-Yao Zhou[a,1], Xiao-Li Xie[a,1,*], Yong-Gang Peng[b], Meng-Jun Wu[c], Xiao-Zhi Deng[a], Ying Wu[d], Li-Jing Xiong[a], Li-Hong Shang[a]

[a] Department of Pediatric Infection and Gastroenterology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China
[b] Department of Anesthesiology, Department of Anesthesiology, University of Florida College of Medicine, Gainesville, Fl, USA
[c] Department of Anesthesiology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China
[d] Department of Pediatric Pneumology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, P.R. China

ARTICLE INFO

Article history:
Received 31 March 2020
Received in revised form 29 April 2020
Accepted 30 April 2020

Keywords:
Coronavirus
SARS
COVID-19
SARS-CoV-2
2019-nCoV
Children

ABSTRACT

Introduction: Coronaviruses, both SARS-CoV and SARS-CoV-2, first appeared in China. They have certain biological, epidemiological and pathological similarities. To date, research has shown that their genes exhibit 79% of identical sequences and the receptor-binding domain structure is also very similar. There has been extensive research performed on SARS; however, the understanding of the pathophysiological impact of coronavirus disease 2019 (COVID-19) is still limited.
Methods: This review drew upon the lessons learnt from SARS, in terms of epidemiology, clinical characteristics and pathogenesis, to further understand the features of COVID-19.
Results: By comparing these two diseases, it found that COVID-19 has quicker and wider transmission, obvious family agglomeration, and higher morbidity and mortality. Newborns, asymptomatic children and normal chest imaging cases emerged in COVID-19 literature. Children starting with gastrointestinal symptoms may progress to severe conditions and newborns whose mothers are infected with COVID-19 could have severe complications. The laboratory test data showed that the percentage of neutrophils and the level of LDH is higher, and the number of CD4+ and CD8+T-cells is decreased in children's COVID-19 cases.
Conclusion: Based on these early observations, as pediatricians, this review put forward some thoughts on children's COVID-19 and gave some recommendations to contain the disease.
© 2020 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

A cluster of patients presented with pneumonia caused by an unknown pathogen that was linked to the seafood wholesale market in Wuhan, China, in December 2019. Subsequently, a new coronavirus was identified by sequencing the whole genome of patient samples (Zhu et al., 2020a). It was named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the Coronavirus Study Group (CSG) of the International Committee on Taxonomy of Viruses (Gorbalenya et al., 2020), and the disease caused by the virus was named coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO).

Of seven coronaviruses identified from humans, HCoV-229E and HCoV-NL63 belong to α-coronaviruses, and HCoV-OC43, MERS-CoV, SARS-CoV and SARS-CoV-2 belong to β-coronaviruses. Both SARS-CoV and SARS-CoV-2 first emerged in China. Although the genome-wide similarity is about 79%, the similarity of the seven conserved domains used for virus identification is as high as 94.6%. This indicates that SARS-CoV-2 belongs to the same genus as SARS-CoV. Additionally, studies have shown that SARS-CoV-2 could enter cells through angiotensin-converting enzyme 2 (ACE2) receptors on the surface of cell membranes, which is consistent with SARS-CoV (Lu et al., 2020a; Zhou et al., 2020).

* Corresponding author. Department of Pediatric Infection and Gastroenterology, Chengdu Women's and Children's Central Hospital, School of Medicine, University of Electronic Science and Technology, Chengdu, Sichuan, 610091, China. Tel.: +86 13438234411.
E-mail address: xxlilye@qq.com (X.-L. Xie).
[1] These two authors contributed equally.

https://doi.org/10.1016/j.ijid.2020.04.090
1201-9712/© 2020 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

⇐

- Trustworthy
- Fact-oriented, precise language
- Slow publication process

- Unreliable
- Emotional language
- Fast publication

⇒

Differences between Scientific Papers and Social Media


International Journal of Infectious Diseases 96 (2020) 710–714

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

From SARS to COVID-19: What we have learned about children infected with COVID-19

Meng-Yao Zhou, Xiao-Li Xie, Yong-Gang Peng, Meng-Jun Wu, Xiao-Zhi Deng, Ying Wu, Li-Jing Xiong, Li-Hong Shang

ARTICLE INFO

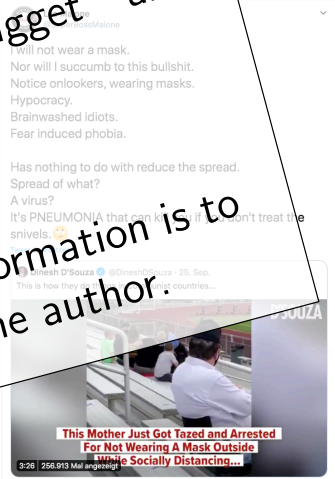
ABSTRACT

1. Introduction

- Trustworthy
- Fact-oriented, precise language
- Slow publication process

⇐ ⇒

- Unreliable
- Emotional language
- Fast publication

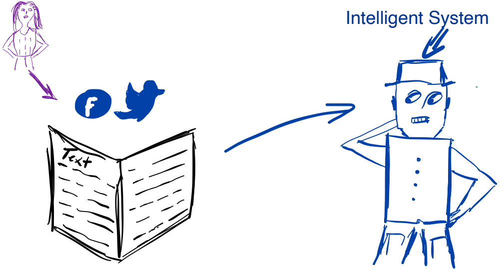


Scientific text: important information nugget are facts, entities, relations.

Social media: an important piece of information is to gain knowledge about the author.

Outline

Overview



What can we learn about the author of a message?

- Personality traits
- Categories (gender, race, nationality, age)
- Expressed emotion, stance, sentiment

Definition of emotions and their linguistic realizations

Emotion (Scherer, 2005)

Definition of emotions and their linguistic realizations

Emotion (Scherer, 2005)

Emotions are "an episode of interrelated, synchronized changes in the states of [...] five organismic subsystems in response to the evaluation of a [...] stimulus-event ..."

Definition of emotions and their linguistic realizations

Emotion (Scherer, 2005)

Emotions are "an episode of interrelated, synchronized changes in the states of [...] five organismic subsystems in response to the evaluation of a [...] stimulus-event ..."

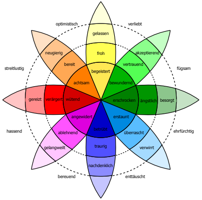
Ekman (1999)

Definition of emotions and their linguistic realizations

Emotion (Scherer, 2005)

Emotions are “an episode of interrelated, synchronized changes in the states of [...] five organismic subsystems in response to the evaluation of a [...] stimulus-event ...”

Ekman (1999)



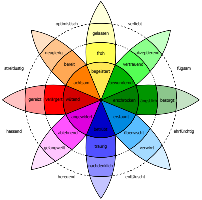
Plutchik (2001)

Definition of emotions and their linguistic realizations

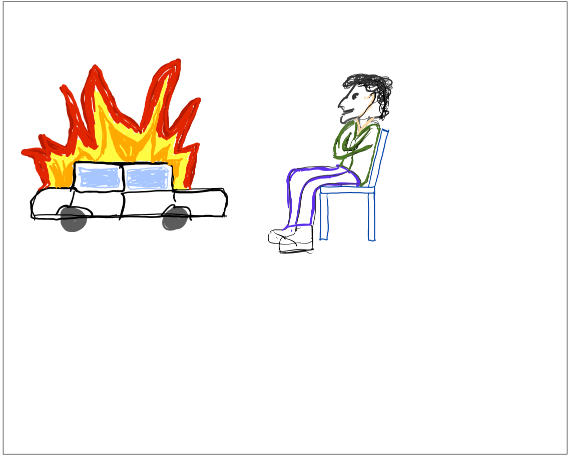
Emotion (Scherer, 2005)

Emotions are "an episode of interrelated, synchronized changes in the states of [...] five organismic subsystems in response to the evaluation of a [...] stimulus-event ..."

Ekman (1999)



Plutchik (2001)

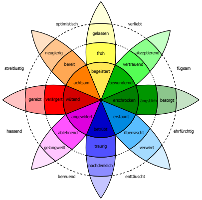


Definition of emotions and their linguistic realizations

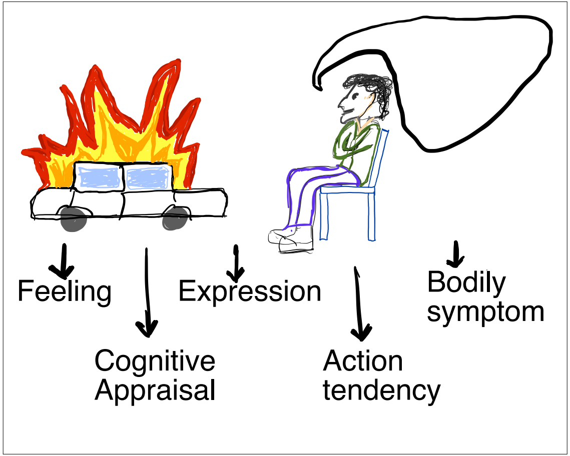
Emotion (Scherer, 2005)

Emotions are "an episode of interrelated, synchronized changes in the states of [...] five organismic subsystems in response to the evaluation of a [...] stimulus-event ..."

Ekman (1999)



Plutchik (2001)

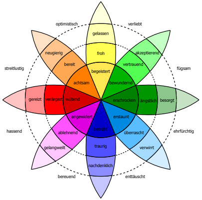


Definition of emotions and their linguistic realizations

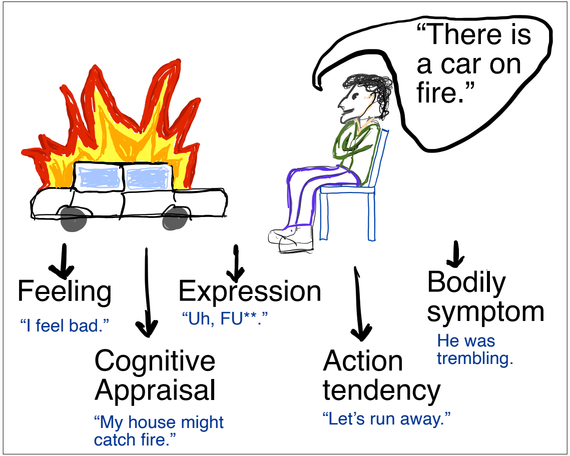
Emotion (Scherer, 2005)

Emotions are "an episode of interrelated, synchronized changes in the states of [...] five organismic subsystems in response to the evaluation of a [...] stimulus-event ..."

Ekman (1999)



Plutchik (2001)



Dictionaries?

- A list of emotion synonyms? No.

Dictionaries?

- A list of emotion synonyms? No.
- Dictionaries exist!

Dictionaries?

- A list of emotion synonyms? No.
- Dictionaries exist!
- Popular Example: NRC Dictionary

Dictionaries?

- A list of emotion synonyms? No.
- Dictionaries exist!
- Popular Example: NRC Dictionary

Examples

- **Anger:**
  aggression, devil, neglected, obstacle

Dictionaries?

- A list of emotion synonyms? No.
- Dictionaries exist!
- Popular Example: NRC Dictionary

Examples

- **Anger:**
  aggression, devil, neglected, obstacle
- **Joy:**
  aesthetics, achieve, cathedral, laughter

Dictionaries?

- A list of emotion synonyms? No.
- Dictionaries exist!
- Popular Example: NRC Dictionary

Examples

- **Anger:**
  aggression, devil, neglected, obstacle
- **Joy:**
  aesthetics, achieve, cathedral, laughter
- **Sadness:**
  unfair, scarce, napkin, tough

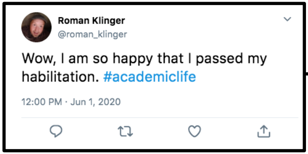
Dictionaries?

- A list of emotion synonyms? No.
- Dictionaries exist!
- Popular Example: NRC Dictionary
- This is a rich resource, performance depends on application and domain.

Examples

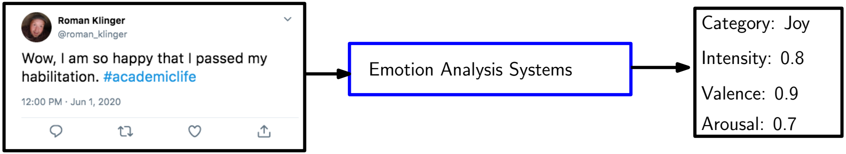
- **Anger:**
  aggression, devil, neglected, obstacle
- **Joy:**
  aesthetics, achieve, cathedral, laughter
- **Sadness:**
  unfair, scarce, napkin, tough

Starting point and Motivation

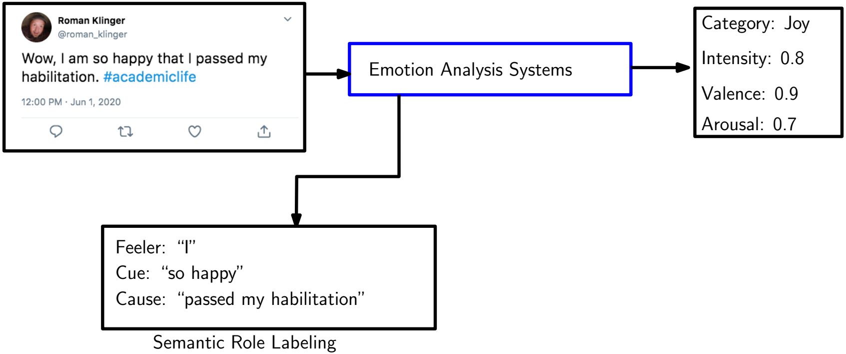


→ Emotion Analysis Systems

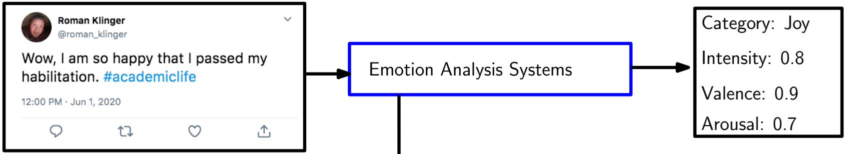
Starting point and Motivation

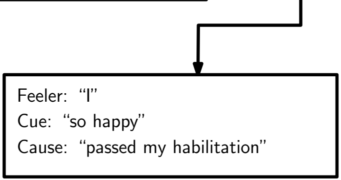


Starting point and Motivation



Starting point and Motivation

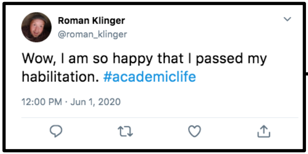




Semantic Role Labeling

- Provide more information to downstream applications

Starting point and Motivation



Emotion Analysis Systems

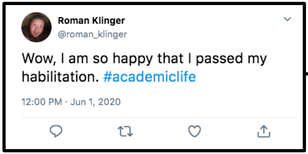
Category: Joy
Intensity: 0.8
Valence: 0.9
Arousal: 0.7

Feeler: “I”
Cue: “so happy”
Cause: “passed my habilitation”

Semantic Role Labeling

- Provide more information to downstream applications
- Challenging sequence labeling task

Starting point and Motivation



Emotion Analysis Systems

Category: Joy
Intensity: 0.8
Valence: 0.9
Arousal: 0.7

Feeler: “I”
Cue: “so happy”
Cause: “passed my habilitation”

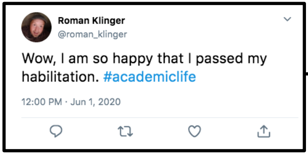
Semantic Role Labeling

- Provide more information to downstream applications
- Challenging sequence labeling task

Who Feels What and Why?
Annotation of a Literature Corpus with Semantic Roles of Emotions

Evgeny Kim and **Roman Klinger**
Institut für Maschinelle Sprachverarbeitung
University of Stuttgart, Pfaffenwaldring 5b, 70569 Stuttgart, Germany
evgeny.kim@ims.uni-stuttgart.de
roman.klinger@ims.uni-stuttgart.de

Starting point and Motivation



Emotion Analysis Systems

Category: Joy
Intensity: 0.8
Valence: 0.9
Arousal: 0.7

Feeler: “I”
Cue: “so happy”
Cause: “passed my habilitation”

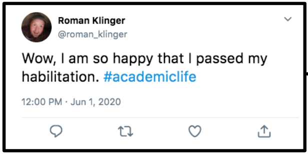
Semantic Role Labeling

- Provide more information to downstream applications
- Challenging sequence labeling task

GoodNewsEveryone: A Corpus of News Headlines Annotated with Emotions, Semantic Roles, and Reader Perception

Laura Bostan, Evgeny Kim, Roman Klinger
Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart
Pfaffenwaldring 5b, 70569 Stuttgart, Germany
{laura.bostan, evgeny.kim, roman.klinger}@ims.uni-stuttgart.de

Starting point and Motivation



Emotion Analysis Systems

Category: Joy
Intensity: 0.8
Valence: 0.9
Arousal: 0.7

Feeler: “I”
Cue: “so happy”
Cause: “passed my habilitation”

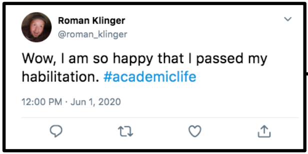
Semantic Role Labeling

- Provide more information to downstream applications
- Challenging sequence labeling task

Token Sequence Labeling vs. Clause Classification for English Emotion Stimulus Detection

Laura Oberländer and **Roman Klinger**
Institut für Maschinelle Sprachverarbeitung, University of Stuttgart
Pfaffenwaldring 5b, 70569 Stuttgart, Germany
{laura.oberlaender,roman.klinger}@ims.uni-stuttgart.de

Starting point and Motivation



→ Emotion Analysis Systems →

Category: Joy
Intensity: 0.8
Valence: 0.9
Arousal: 0.7

Feeler: "I"
Cue: "so happy"
Cause: "passed my habilitation"

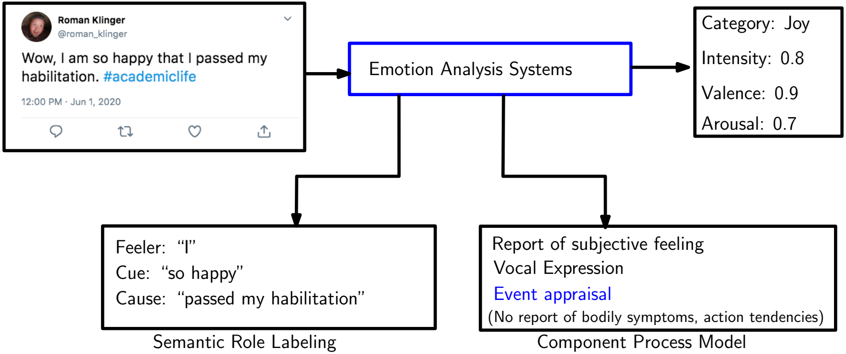
Semantic Role Labeling

- Provide more information to downstream applications
- Challenging sequence labeling task

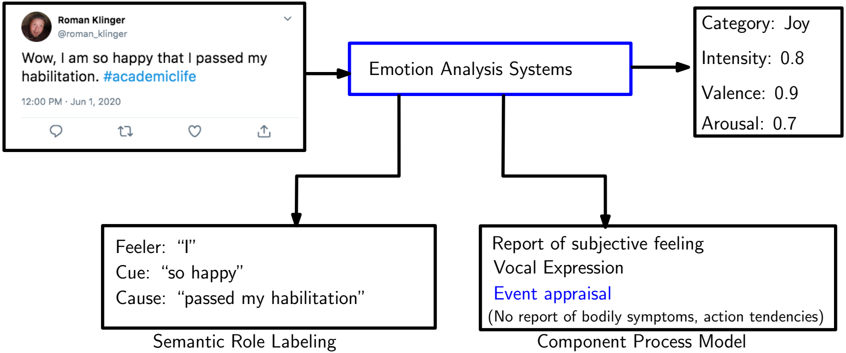
Experiencers, Stimuli, or Targets:
Which Semantic Roles Enable Machine Learning to Infer the Emotions?

Laura Oberländer, Kevin Reich, and **Roman Klinger**
Institut für Maschinelle Sprachverarbeitung, University of Stuttgart, Germany
`{firstname.lastname}@ims.uni-stuttgart.de`

Starting point and Motivation

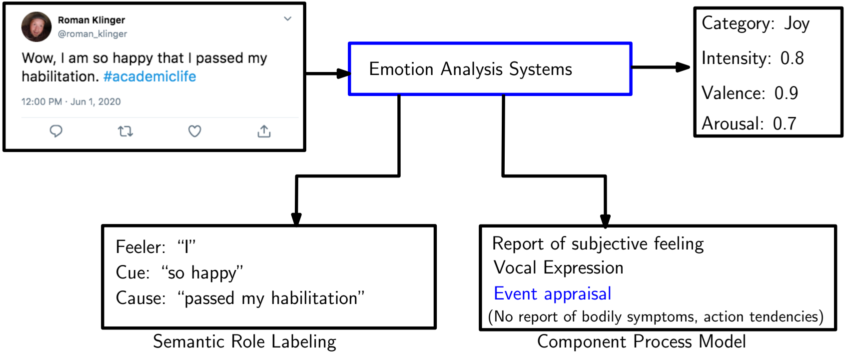


Starting point and Motivation



- Additional information for downstream applications

Starting point and Motivation



- Additional information for downstream applications
- Supports emotion detection

Intermediate Results on Emotion Component Model

Intermediate Results on Emotion Component Model

- Is this really the case?

Intermediate Results on Emotion Component Model

- Is this really the case?
- All components play a role in each emotion?

Intermediate Results on Emotion Component Model

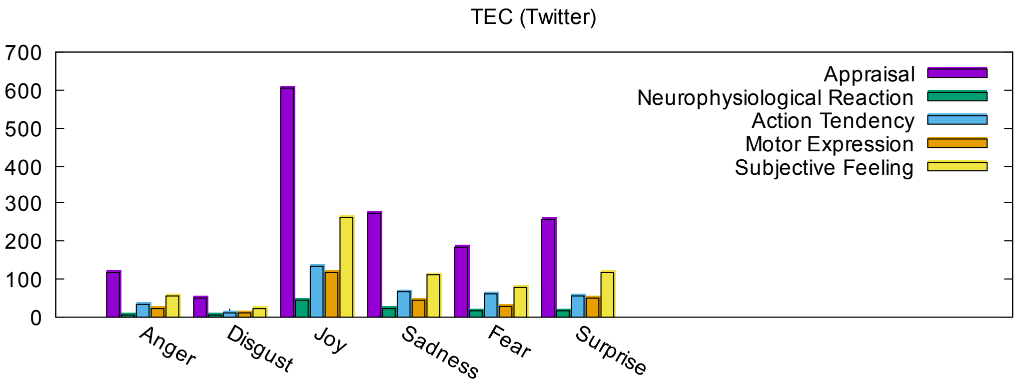
- Is this really the case?
- All components play a role in each emotion?
- How can recognizing the components contribute to emotion recognition then?

Intermediate Results on Emotion Component Model

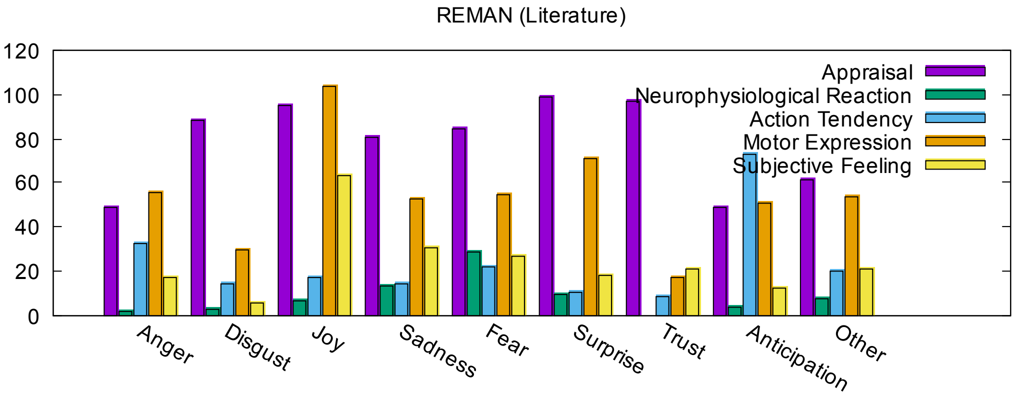
- Is this really the case?
- All components play a role in each emotion?
- How can recognizing the components contribute to emotion recognition then?

⇒ Annotation study on literature and Twitter
(part of the recent theses by Amelie Heindl and Felix Casel)

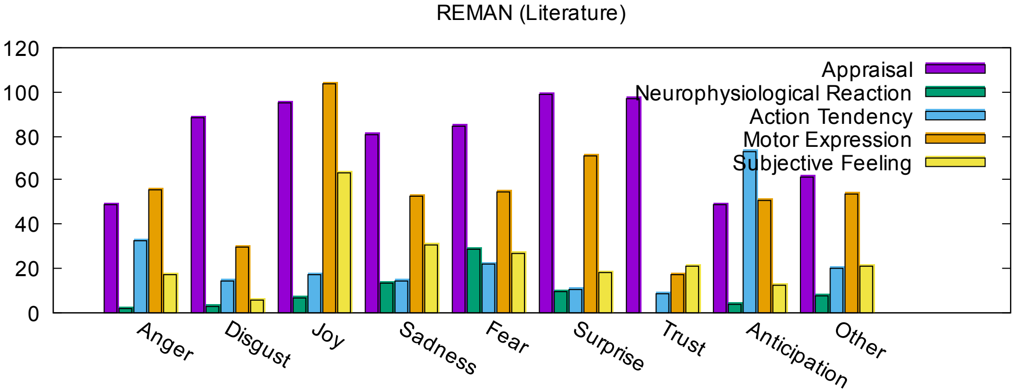
Intermediate Results on Emotion Component Model



Intermediate Results on Emotion Component Model

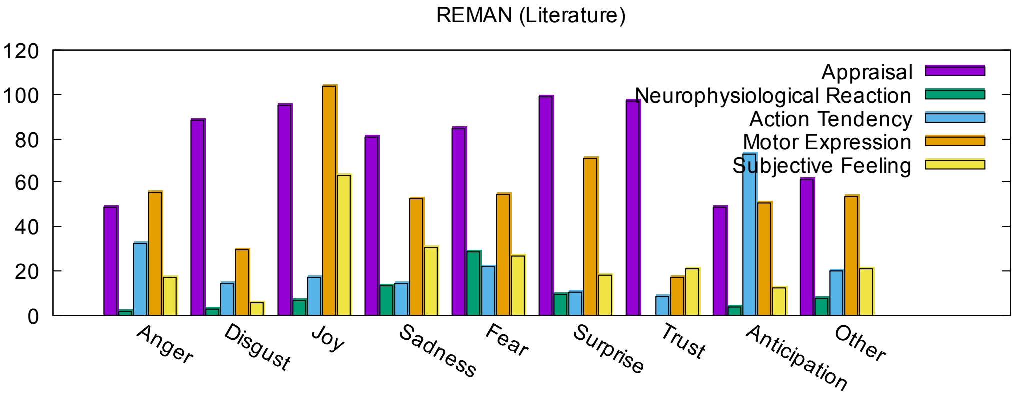


Intermediate Results on Emotion Component Model



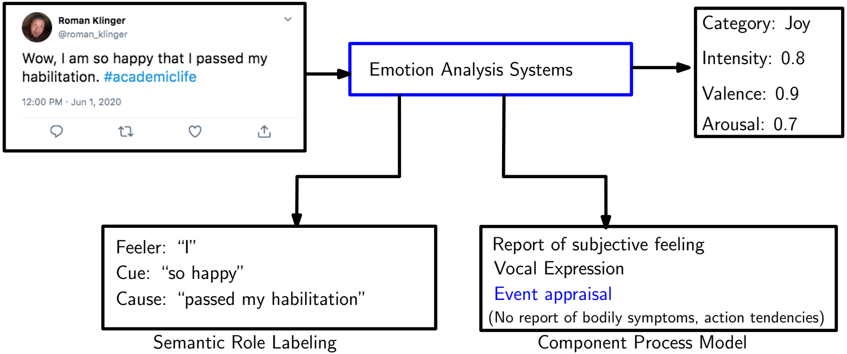
- Providing component information to emotion classifier helps in literature

Intermediate Results on Emotion Component Model

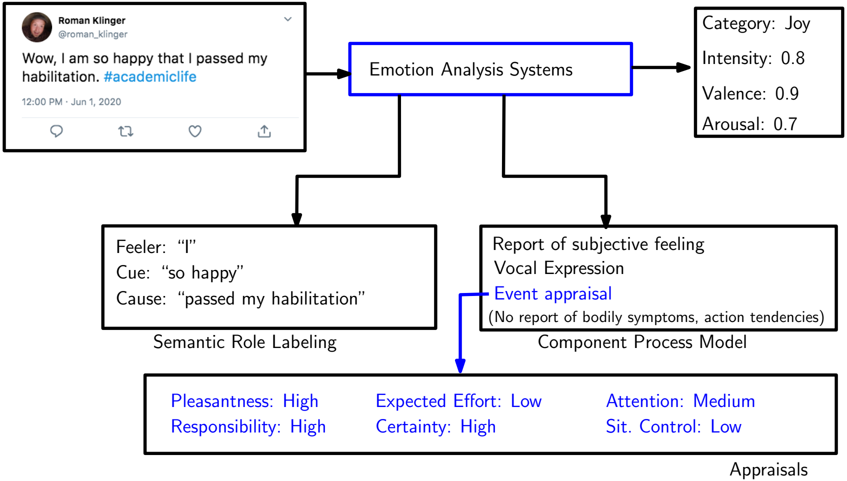


- Providing component information to emotion classifier helps in literature
- Multi-task learning of components and emotions shows improvements for both corpora

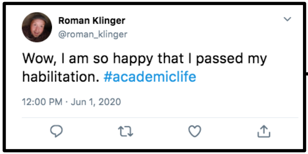
Starting point and Motivation



Starting point and Motivation



Starting point and Motivation



Emotion Analysis Systems

Category: Joy
Intensity: 0.8
Valence: 0.9
Arousal: 0.7

Feeler: “I”
Cue: “so happy”
Cause: “passed my habilitation”

Semantic Role Labeling

Report of subjective feeling
Vocal Expression
Event appraisal
(No report of bodily symptoms, action tendencies)

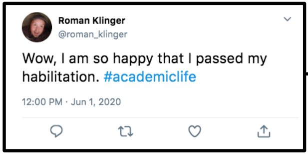
Component Process Model

Pleasantness: High
Responsibility: High
Expected Effort: Low
Certainty: High
Attention: Medium
Sit. Control: Low

Appraisals

Category: Joy

Starting point and Motivation



Emotion Analysis Systems

Category: Joy
Intensity: 0.8
Valence: 0.9
Arousal: 0.7

Feeler: "I"
Cue: "so happy"
Cause: "passed my habilitation"

Semantic Role Labeling

Report of subjective feeling
Vocal Expression
Event appraisal
(No report of bodily symptoms, action tendencies)

Pleasantness: High
Responsibility: High
Expected
Certainty:

Category:

Appraisal Theories for Emotion Classification in Text

Jan Hofmann[1], **Enrica Troiano**[1], **Kai Sassenberg**[2,3], and **Roman Klinger**[1]
[1]Institut für Maschinelle Sprachverarbeitung, University of Stuttgart, Germany
[2]Leibniz-Institut für Wissensmedien, Tübingen, Germany
[3]University of Tübingen, Germany
{jan.hofmann,enrica.troiano,roman.klinger}@ims.uni-stuttgart.de
k.sassenberg@iwm-tuebingen.de

Appraisal Annotation

Appraisal Annotation

Most probably, at the time when the event happened, the writer...

Appraisal Annotation

Most probably, at the time when the event happened, the writer…

- …wanted to devote further attention to the event. (Attention)

Appraisal Annotation

Most probably, at the time when the event happened, the writer...

- ...wanted to devote further attention to the event. (Attention)
- ...was certain about what was happening. (Certainty)

Appraisal Annotation

Most probably, at the time when the event happened, the writer...

- ...wanted to devote further attention to the event. (Attention)
- ...was certain about what was happening. (Certainty)
- ...had to expend mental or physical effort to deal with the situation. (Effort)

Appraisal Annotation

Most probably, at the time when the event happened, the writer…

- …wanted to devote further attention to the event. (Attention)
- …was certain about what was happening. (Certainty)
- …had to expend mental or physical effort to deal with the situation. (Effort)
- …found that the event was pleasant. (Pleasantness)

Appraisal Annotation

Most probably, at the time when the event happened, the writer...

- ...wanted to devote further attention to the event. (Attention)
- ...was certain about what was happening. (Certainty)
- ...had to expend mental or physical effort to deal with the situation. (Effort)
- ...found that the event was pleasant. (Pleasantness)
- ...was responsible for the situation. (Responsibility)

Appraisal Annotation

Most probably, at the time when the event happened, the writer...

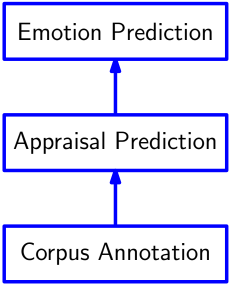
- ...wanted to devote further attention to the event. (Attention)
- ...was certain about what was happening. (Certainty)
- ...had to expend mental or physical effort to deal with the situation. (Effort)
- ...found that the event was pleasant. (Pleasantness)
- ...was responsible for the situation. (Responsibility)
- ...found that he/she was in control of the situation. (Control)

Appraisal Annotation

Most probably, at the time when the event happened, the writer...

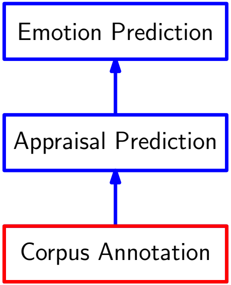
- ...wanted to devote further attention to the event. (Attention)
- ...was certain about what was happening. (Certainty)
- ...had to expend mental or physical effort to deal with the situation. (Effort)
- ...found that the event was pleasant. (Pleasantness)
- ...was responsible for the situation. (Responsibility)
- ...found that he/she was in control of the situation. (Control)
- ...found that the event could not have been changed/influenced by anyone. (Circumstance)

Corpus Selection



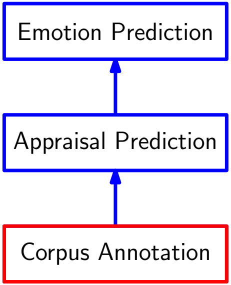
Corpus Selection

- Which corpus to use to study appraisals in text?



Corpus Selection

- Which corpus to use to study appraisals in text?



Crowdsourcing and Validating
Event-focused Emotion Corpora for German and English

Enrica Troiano, Sebastian Padó and **Roman Klinger**
Institut für Maschinelle Sprachverarbeitung
University of Stuttgart, Germany
{firstname.lastname}@ims.uni-stuttgart.de

Corpus Selection

Emotion Prediction

Appraisal Prediction

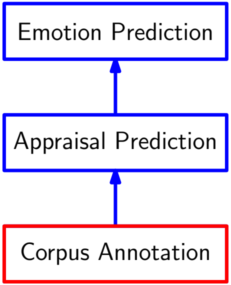
Corpus Annotation

- Which corpus to use to study appraisals in text?
  - "Remember an event which triggered [emotion] and describe it: 'I felt [emotion word], when…' "

Crowdsourcing and Validating
Event-focused Emotion Corpora for German and English

Enrica Troiano, Sebastian Padó and **Roman Klinger**
Institut für Maschinelle Sprachverarbeitung
University of Stuttgart, Germany
{firstname.lastname}@ims.uni-stuttgart.de

Corpus Selection

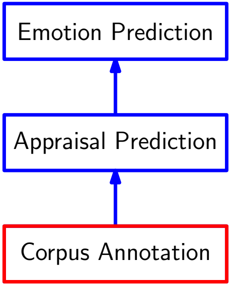


- Which corpus to use to study appraisals in text?
  - "Remember an event which triggered [emotion] and describe it: 'I felt [emotion word], when...' "
  - 1001 event descriptions, stratified by emotion (anger, disgust, fear, guilt, joy, shame, sadness)

Crowdsourcing and Validating
Event-focused Emotion Corpora for German and English

Enrica Troiano, Sebastian Padó and **Roman Klinger**
Institut für Maschinelle Sprachverarbeitung
University of Stuttgart, Germany
{firstname.lastname}@ims.uni-stuttgart.de

Corpus Selection

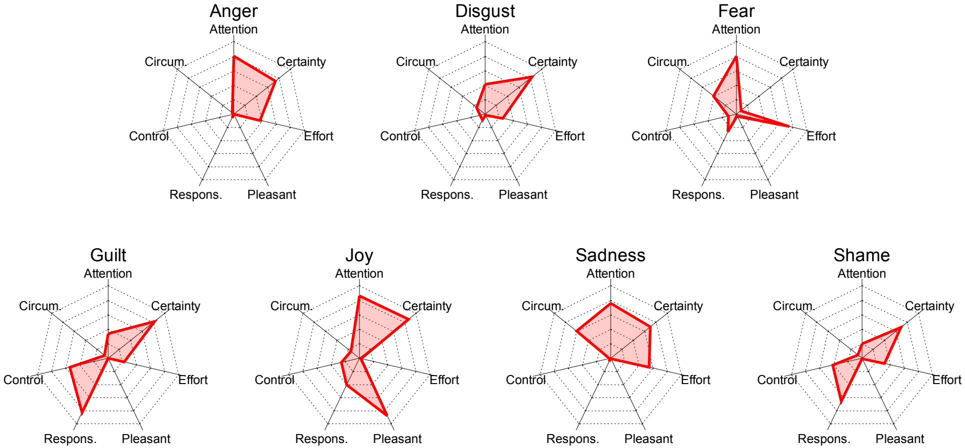


- Which corpus to use to study appraisals in text?
  - "Remember an event which triggered [emotion] and describe it: 'I felt [emotion word], when...' "
  - 1001 event descriptions, stratified by emotion (anger, disgust, fear, guilt, joy, shame, sadness)

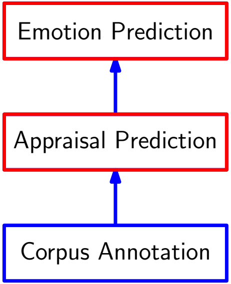
Examples

- I felt [sadness] when I saw a homeless cat on the street.
- I felt [shame] when someone commented that I was looking very untidy.
- I felt [anger] when the police did not update me on a crime.

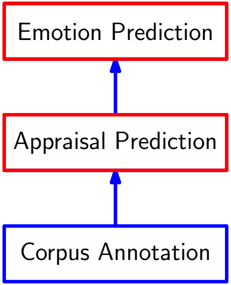
Annotation Results

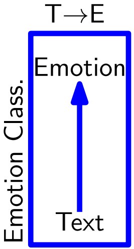


Modelling and Experimental Setting

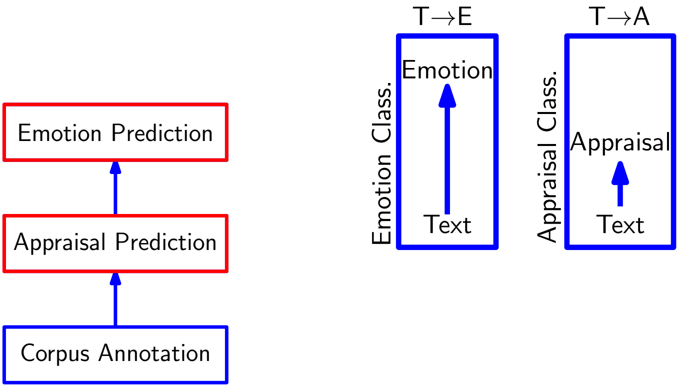


Modelling and Experimental Setting

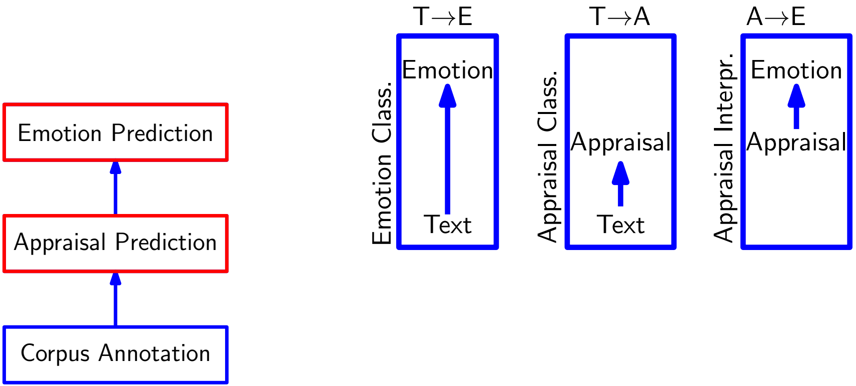




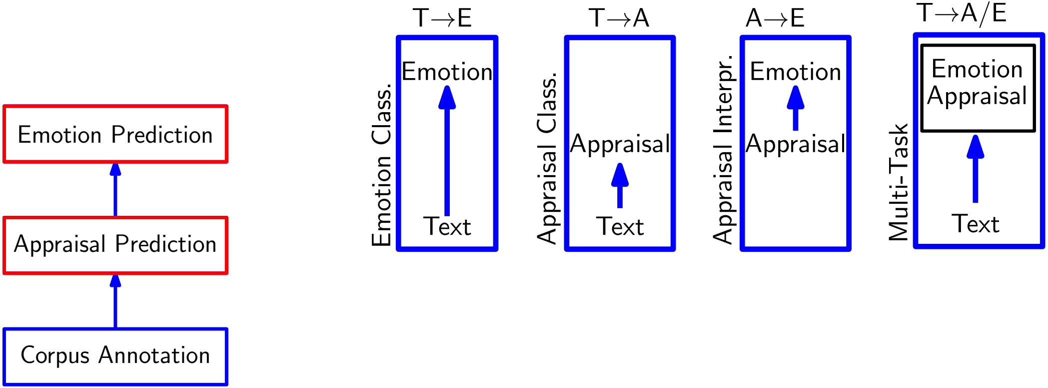
Modelling and Experimental Setting



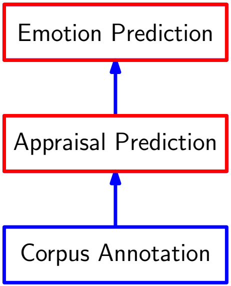
Modelling and Experimental Setting

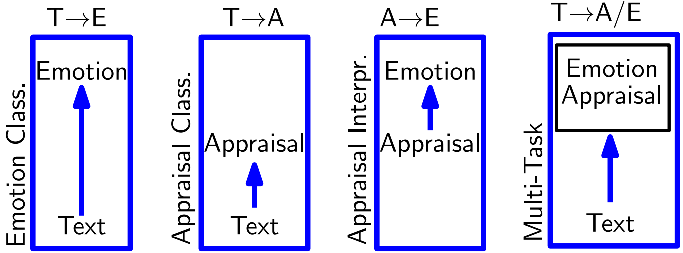


Modelling and Experimental Setting



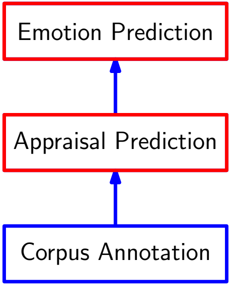
Modelling and Experimental Setting

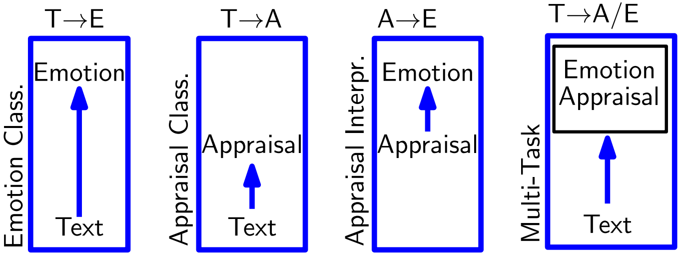




- Text classifiers: Convolutional neural network, pretrained GloVe 300-dimensional embeddings, filter sizes 2,3,4, ReLu activation, dropout 0.5

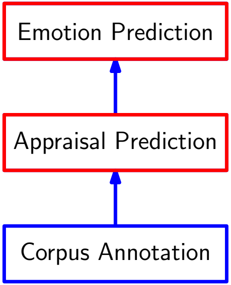
Modelling and Experimental Setting

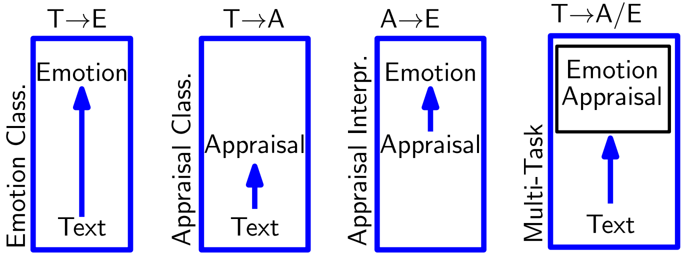




- Text classifiers: Convolutional neural network, pretrained GloVe 300-dimensional embeddings, filter sizes 2,3,4, ReLu activation, dropout 0.5
- Emotion from Appraisal:
  Fully connected neural network with two layers

Modelling and Experimental Setting





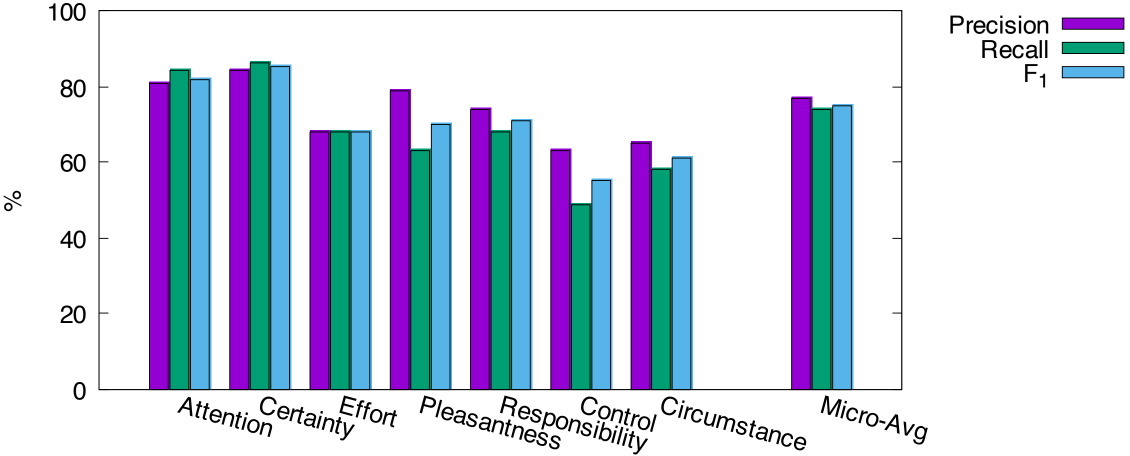
- Text classifiers: Convolutional neural network, pretrained GloVe 300-dimensional embeddings, filter sizes 2,3,4, ReLu activation, dropout 0.5
- Emotion from Appraisal:
  Fully connected neural network with two layers
- Evaluation via 10×10-fold CV

Modelling Results

How well can we predict appraisal dimensions from text?

Modelling Results

How well can we predict appraisal dimensions from text?

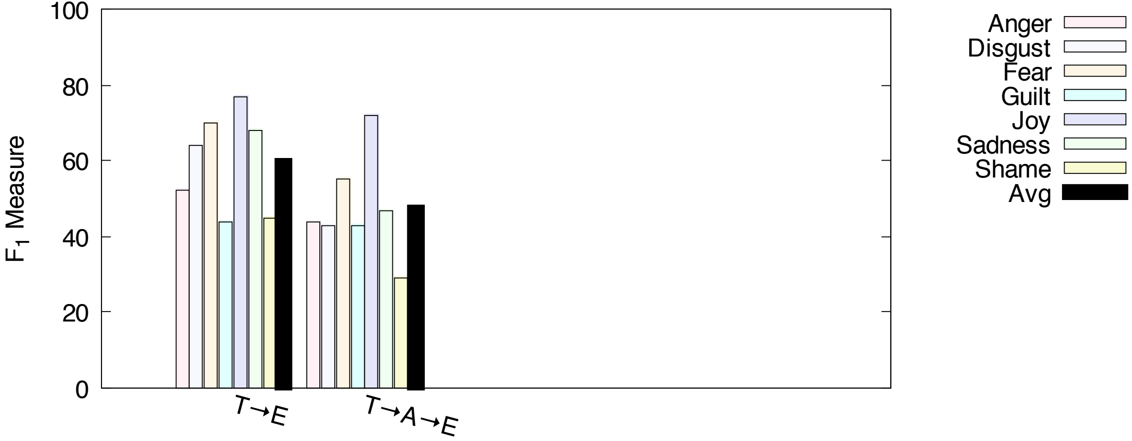


Modelling Results

Can this approach improve emotion classification?

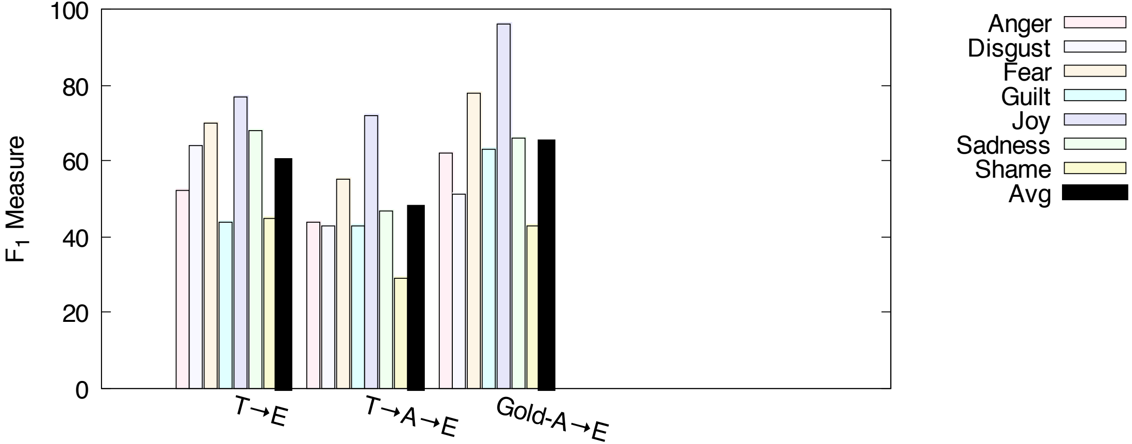
Modelling Results

Can this approach improve emotion classification?



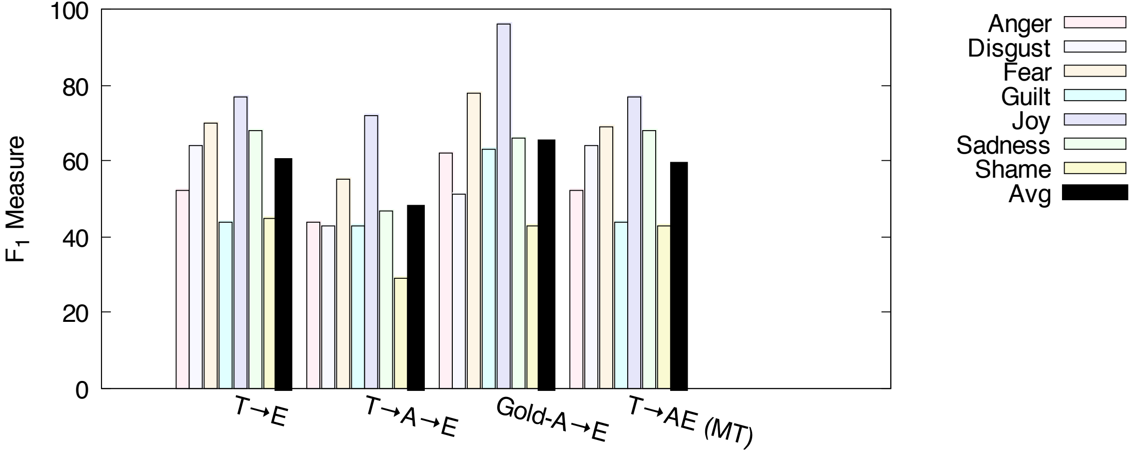
Modelling Results

Can this approach improve emotion classification?



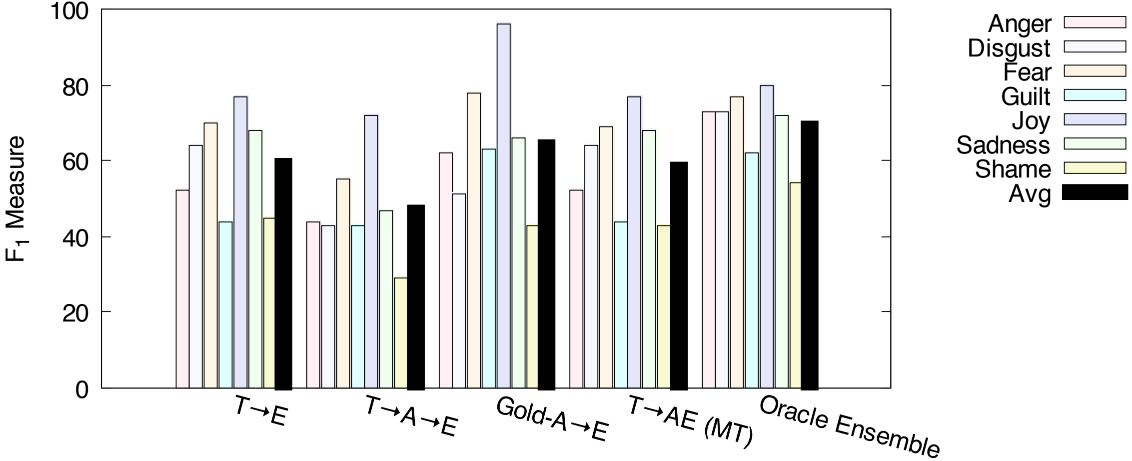
Modelling Results

Can this approach improve emotion classification?



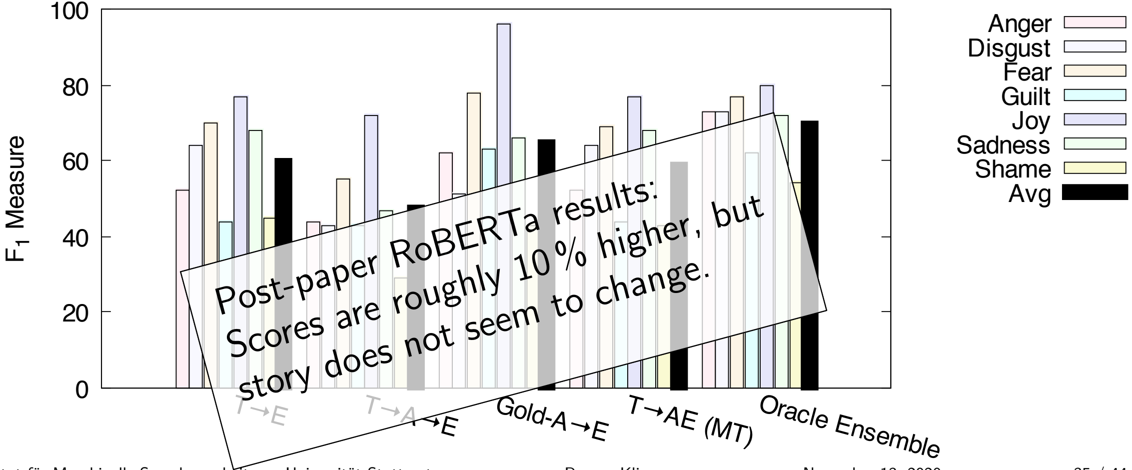
Modelling Results

Can this approach improve emotion classification?



Modelling Results

Can this approach improve emotion classification?



Examples

Gold	A→E	T→E	Text
Anger	Anger	Disgust	when I saw someone mistreating an animal.
Disgust	Disgust	Shame	because I ate a sausage that was horrible.
Disgust	Disgust	Fear	when I was on a ferry in a storm and lots of people were vomiting.
Guilt	Guilt	Shame	when I took something without paying.
Guilt	Guilt	Joy	for denying to offer my kids what they demanded of me.
Joy	Joy	Disgust	when I found a twenty pound note on the ground outside.

Summary

Summary

- Emotion classification directly from text remains the best approach (so far)

Summary

- Emotion classification directly from text remains the best approach (so far)
- Appraisal prediction has potential to improve emotion classification

Summary

- Emotion classification directly from text remains the best approach (so far)
- Appraisal prediction has potential to improve emotion classification
- Oracle approach shows that the two methods are complementary

Summary

- Emotion classification directly from text remains the best approach (so far)
- Appraisal prediction has potential to improve emotion classification
- Oracle approach shows that the two methods are complementary

- Few concepts: possible to tackle as standard text classification approach (optionally enriched and modelled jointly with text segments)

Next (concrete) Steps in Emotion Analysis

Next (concrete) Steps in Emotion Analysis

Appraisal:

- Large scale crowdsourcing of events with appraisal labels from experiencer

Next (concrete) Steps in Emotion Analysis

Appraisal:

- Large scale crowdsourcing of events with appraisal labels from experiencer
- Annotation of structured event representations with appraisals

Next (concrete) Steps in Emotion Analysis

Appraisal:

- Large scale crowdsourcing of events with appraisal labels from experiencer
- Annotation of structured event representations with appraisals
- End-to-end joint pipeline learning of appraisal and emotions

Next (concrete) Steps in Emotion Analysis

Appraisal:

- Large scale crowdsourcing of events with appraisal labels from experiencer
- Annotation of structured event representations with appraisals
- End-to-end joint pipeline learning of appraisal and emotions
- Explore other model architectures

Next (concrete) Steps in Emotion Analysis

Appraisal:

- Large scale crowdsourcing of events with appraisal labels from experiencer
- Annotation of structured event representations with appraisals
- End-to-end joint pipeline learning of appraisal and emotions
- Explore other model architectures
- DFG Project CEAT starts in January 2021 (with Laura Oberländer as postdoc)

Next (concrete) Steps in Emotion Analysis

Appraisal:

- Large scale crowdsourcing of events with appraisal labels from experiencer
- Annotation of structured event representations with appraisals
- End-to-end joint pipeline learning of appraisal and emotions
- Explore other model architectures
- DFG Project CEAT starts in January 2021 (with Laura Oberländer as postdoc)

Role labeling:

Next (concrete) Steps in Emotion Analysis

Appraisal:

- Large scale crowdsourcing of events with appraisal labels from experiencer
- Annotation of structured event representations with appraisals
- End-to-end joint pipeline learning of appraisal and emotions
- Explore other model architectures
- DFG Project CEAT starts in January 2021 (with Laura Oberländer as postdoc)

Role labeling:

- Multimodality: Emotion stimuli in images (DFG project in preparation with C. Silberer)

Next (concrete) Steps in Emotion Analysis

Appraisal:

- Large scale crowdsourcing of events with appraisal labels from experiencer
- Annotation of structured event representations with appraisals
- End-to-end joint pipeline learning of appraisal and emotions
- Explore other model architectures
- DFG Project CEAT starts in January 2021 (with Laura Oberländer as postdoc)

Role labeling:

- Multimodality: Emotion stimuli in images (DFG project in preparation with C. Silberer)
- Joint modelling of roles (final WP in DFG Project SEAT)

Outline

Biomedical entities and emotions

Biomedical entities and emotions

Biomedical information extraction

- Many entities
- Comparably few linguistic realizations
- Huge established databases available
- Precision for finding and linking entities is a challenge

Biomedical entities and emotions

Biomedical information extraction

- Many entities
- Comparably few linguistic realizations
- Huge established databases available
- Precision for finding and linking entities is a challenge

Emotion analysis

- Few concepts to link
- Many linguistic realizations
- Conceptualization under constant discussion
- Precision and recall are both challenging

Why do I care about both?

Why do I care about both?

- Emotions as a fact:
  Which events influence public health, well-being, and quality of life of individuums and peoples?

Why do I care about both?

- Emotions as a fact:
  Which events influence public health, well-being, and quality of life of individuums and peoples?
- Emotions as factor:
  Is misinformation/desinformation correlated to particular expressed emotions?

Why do I care about both?

- Emotions as a fact:
  Which events influence public health, well-being, and quality of life of individuums and peoples?
- Emotions as factor:
  Is misinformation/desinformation correlated to particular expressed emotions?

- New DFG Project starting 2021 on biomedical fact checking (with Amelie Wührl as PhD student)

Wischen and Mischen

Vision and Mission

- Make accessible and understand

Vision and Mission

- Make accessible and understand
  - communication of relational (biomedical) information across different sources, scientific text, social media, experts and laypeople

Vision and Mission

- Make accessible and understand
  - communication of relational (biomedical) information across different sources, scientific text, social media, experts and laypeople
  - realizations of psychological concepts like emotions across different domains, modalities and realization patterns

Vision and Mission

- Make accessible and understand
  - communication of relational (biomedical) information across different sources, scientific text, social media, experts and laypeople
  - realizations of psychological concepts like emotions across different domains, modalities and realization patterns
- Link medical information and psychological concepts as they occur “in the wild”.

Vision and Mission

- Make accessible and understand
  - communication of relational (biomedical) information across different sources, scientific text, social media, experts and laypeople
  - realizations of psychological concepts like emotions across different domains, modalities and realization patterns
- Link medical information and psychological concepts as they occur "in the wild".
  - Downstream tasks: fact-checking, misinformation detection, pharmacovigilance, opinion mining, ...
- Develop resources and machine learning methods to enable these goals.

Ethical Considerations

Ethical Considerations

- Systems suffer from biases

Ethical Considerations

- Systems suffer from biases
- Systems are not reliable

Ethical Considerations

- Systems suffer from biases
- Systems are not reliable
- Corpora and systems do not represent all groups in a population equally

Ethical Considerations

- Systems suffer from biases
- Systems are not reliable
- Corpora and systems do not represent all groups in a population equally
- Concepts are analyzed which people might not even be aware of

Ethical Considerations

- Systems suffer from biases
- Systems are not reliable
- Corpora and systems do not represent all groups in a population equally
- Concepts are analyzed which people might not even be aware of
- Analyses should never enable any inference about individuals, results should only be reported in aggregated form.

Thank you for your attention.
Questions? Remarks?

?

(please type a Q in the chat if you have a question)

Universität Stuttgart
Institut für
Maschinelle Sprachverarbeitung

Computational Natural Language Understanding: Use cases in the life sciences and psychology

Inaugural Lecture

November 13, 2020

Roman Klinger
roman.klinger@ims.uni-stuttgart.de

@roman_klinger romanklinger
http://www.romanklinger.de/