Sentiment Analysis and Opinion Mining

Prof. Alfio Ferrara

**Department of Computer Science, Università degli Studi di Milano
Room 7012 via Celoria 18, 20133 Milano, Italia alfio.ferrara@unimi.it**

sed noli modo

*"The only thing worse than being talked about,
is not being talked about."
— Oscar Wilde.*

Sentiment analysis and opinion mining is the field of study that analyzes people's **opinions, sentiments, evaluations, attitudes, and emotions** from **written language**.

It is one of the most active research areas in **natural language processing** and is also widely studied in **data mining**, **Web mining**, and **text mining**.

Liu, B. (2012). *Sentiment analysis and opinion mining.* Synthesis lectures on human language technologies, 5(1), 1-167.

Levels of Analysis

Document level

To classify whether a whole opinion document expresses a positive or negative sentiment.
This level of analysis assumes that each document expresses opinions on a single entity.

Sentence level

To determine whether each sentence expressed a positive, negative, or neutral opinion.
This level of analysis is closely related to *subjectivity* classification in that it tries to distinguish between *objective* and *subjective* sentences.

Entity level

To directly detect opinions. An opinion consists of a sentiment (positive or negative) and a target (of opinion).
For example, the sentence "*the iPhone's call quality is good, but its battery life is short.*" evaluates two aspects: call quality (positive) and battery life (negative), of iPhone (entity).

Levels of Analysis (example)

	DOCUMENT	SENTENCE	ENTITY
Rented in Phoenix.		**N**	**N** → *Phoenix*
Last minute rental so getting into a car with a cigarette burn on the drivers seat told me this was not a non smoking car.	**+**	**-**	**-** → *car*
The Good counter lady was nice, polite and efficient, so that, thanks to Phoenix I could put captain Bars on my sons Army uniform.		**+**	**+** → *counter lady* **+** → *Phoenix*

Sentiment Lexicon

One of the most straightforward ways of addressing sentiment analysis is to define a **lexicon of sentiment words** or **opinion words** (or even small key phrases) that may be used as indicators for sentiment.

- Baccianella, S., Esuli, A., & Sebastiani, F. (2010, May). Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In Lrec (Vol. 10, No. 2010, pp. 2200-2204). https://github.com/aesuli/sentiwordnet

- Musto, C., Semeraro, G., & Polignano, M. (2014, December). A Comparison of Lexicon-based Approaches for Sentiment Analysis of Microblog Posts. In DART@ AI* IA (pp. 59-68).

- De Albornoz, J. C., Plaza, L., & Gervás, P. (2012, May). SentiSense: An easily scalable concept-based affective lexicon for sentiment analysis. In LREC (Vol. 12, pp. 3562-3567).

Issues with Sentiment Lexicon (I)

*Word orientation depends on the **application context** and **word sense disambiguation***

Issues with Sentiment Lexicon (II)

A sentence containing sentiment words may not express any sentiment

Interrogative sentences pretty often do not express sentiment

can you tell me which Sony camera is good ?

… but sometimes they do

Does anyone know how to repair this terrible printer ?

Also conditional sentences pretty often do not express sentiment

if i can find a good camera in the shop, i will buy it

… but sometimes they do

if you are looking for a good car, get toyota camry.

Issues with Sentiment Lexicon (III)

"... no dark sarcasm in the classroom ..."

*What a **great** car! it stopped working in two days.*

Issues with Sentiment Lexicon (IV)

Many (objective) sentences without sentiment words can also imply opinions

this fridge uses a lot of energy

after sleeping on the mattress for two days, a valley has formed in the middle
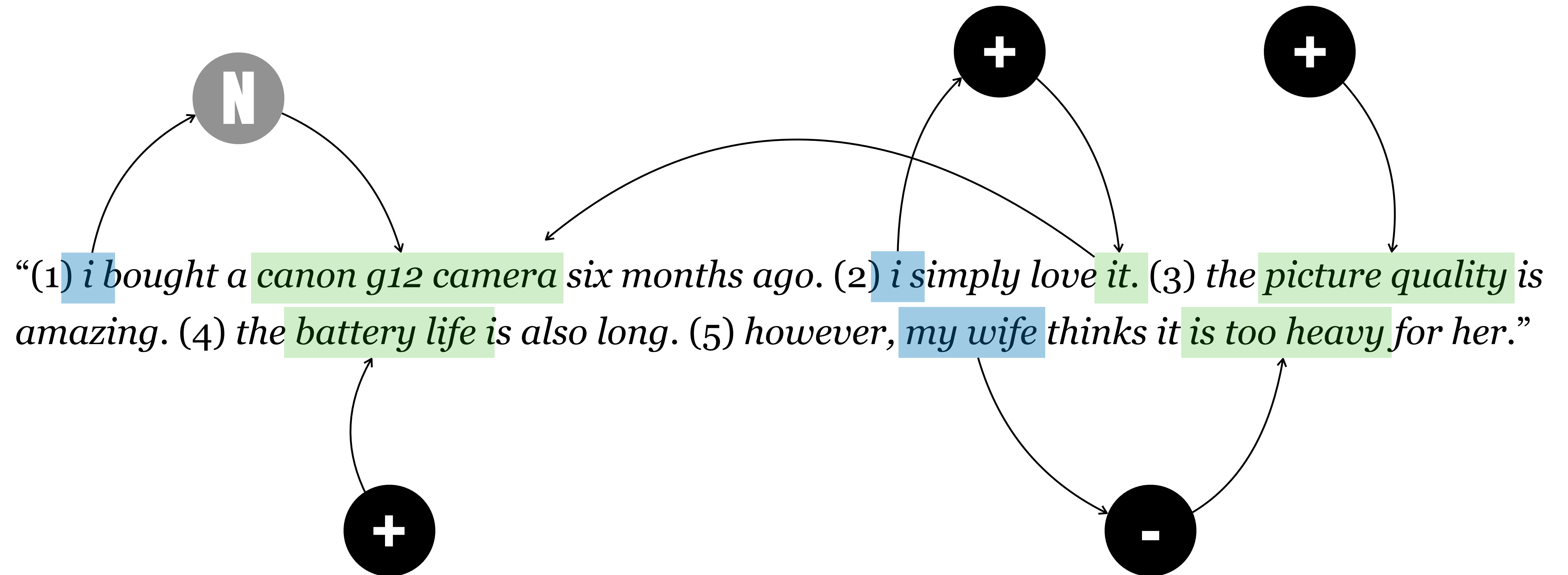
Problems and Definitions in Sentiment Analysis

"(1) i bought a canon g12 camera six months ago. (2) i simply love it. (3) the picture quality is amazing. (4) the battery life is also long. (5) however, my wife thinks it is too heavy for her."

Opinion definition (I)

Definition (Opinion): An opinion is a quadruple, (g, s, h, t), where **g** is the opinion (or sentiment) **target**, **s** is the sentiment about the target (either polarity or score or even an emotion), **h** is the **opinion holder**, and **t** is the **time** when the opinion was expressed.

Entity definition (I)

Definition (entity): An entity *e* is a product, service, topic, issue, person, organization, or event. It is described with a pair, *e: (t, W)*, where **t** is a **hierarchy of parts**, sub-parts, and so on, and **W** is a set of **attributes** of e. Each part or sub-part also has its own set of attributes.

Entity example

A particular model of camera is an entity, e.g., Canon G12. It has a set of attributes, e.g., picture quality, size, and weight, and a set of parts, e.g., lens, viewfinder, and battery. Battery also has its own set of attributes, e.g., battery life and battery weight.

A topic can be an entity too, e.g., tax increase, with its parts "tax increase for the poor," "tax increase for the middle class," and "tax increase for the rich."

Canon G12 10 MP Digital Camera with 5x Optical Image Stabilized Zoom and 2.8 Inch Vari-Angle LCD
by Canon

⭐⭐⭐⭐✫ | 482 ratings | 31 answered questions

Available from these sellers.

Style: **US Version**

- 10.0-megapixel sensor and the DIGIC 4 Image Processor combine to create Canon's HS SYSTEM for improved low light performance
- Shoot 720p HD video in stereo sound; HDMI output
- Canon's Hybrid IS compensates for angular and shift camera shake during close-up shooting
- 5x optical zoom with Optical Image Stabilizer; 28mm wide-angle lens; optical viewfinder
- Capture images and video to SD/SDHC/SDXC memory card, MultiMediaCard, MMC Plus Card, HC MMC Plus Card (not included)

Used (19) from $120.00

Opinion definition (II)

Definition (opinion): An opinion is a quintuple, (e, a, s_{ea}, h, t), where e is an **entity**, a is an aspect of e, s_{ea} is the **sentiment** on aspect a of entity e, h is the **opinion holder**, and t is the **time** when the opinion is expressed by h. The sentiment s_{ea} is positive, negative, or neutral, or expressed with different strength /intensity levels, e.g., 1–5 stars as used by most review sits on the Web.
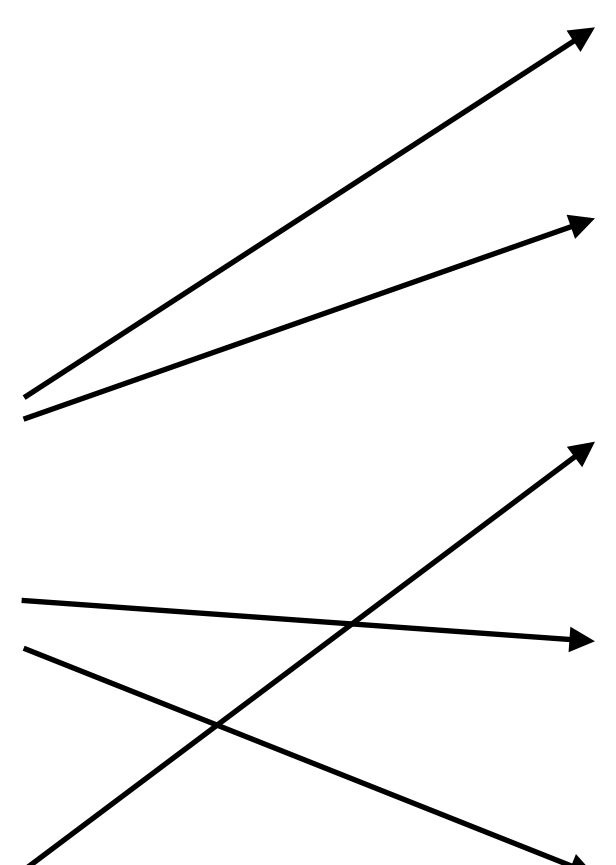
Objective of sentiment analysis: Given an opinion document d, discover all opinion quintuples (e, a, s_{ea}, h, t) in d.

Named Entity Recognition (Ner)	→	Aspect Extraction	→	Opinion Holder Extraction	→	Time Extraction	→	Aspect Sentiment Classification

Example: Posted by: big John Date: Sept. 15, 2011

(1) *i bought a Samsung camera and my friends brought a canon camera yesterday.* (2) *in the past week, we both used the cameras a lot.* (3) *the photos from my Samy are not that great, and the battery life is short too.* (4) *my friend was very happy with his camera and loves its picture quality.* (5) *i want a camera that can take good photos.* (6) *i am going to return it tomorrow.*

1	Samsung camera	Photo	-	Big John	15.9.2011
2	Samsung camera	Battery life	-	Big John	15.9.2011
3	Samsung camera	General	-	Big John	15.9.2011
4	Canon camera	General	+	Big John friend	15.9.2011
5	Canon camera	Photo	+	Big John friend	15.9.2011

| Named Entity Recognition (Ner) | Aspect Extraction | Opinion Holder Extraction | Time Extraction | Aspect Sentiment Classification |

Types of opinion

Opinion	Regular	Comparative
Explicit	*"coke tastes great,"*	*"coke tastes better than Pepsi."*
Implicit	*"i bought the mattress a week ago, and a valley has formed,"*	*"the battery life of nokia phones is longer than Samsung phones."*

Subjectivity

Definition (sentence subjectivity): An *objective sentence* presents some *factual information* about the world, while *a subjective sentence* expresses some *personal feelings, views, or beliefs.*

- **A subjective sentence may not express any sentiment.** For example, *"I think that he went home"* is a subjective sentence, but does not express any sentiment, as well as *"I want a camera that can take good photos"*.
- **Objective sentences can imply opinions or sentiments due to desirable and undesirable facts**. For example, the following two sentences which state some facts clearly imply negative sentiments (which are implicit opinions) about their respective products because the facts are undesirable:

"the earphone broke in two days."
"i brought the mattress a week ago and a valley has formed "

Emotions

Definition (emotion): Emotions are our subjective feelings and thoughts.

- **Rational evaluation:** Such evaluations are from rational reasoning, tangible beliefs, and utilitarian attitudes. For example, the following sentences express rational evaluations:

    "the voice of this phone is clear,"

    "this car is worth the price," and

    "i am happy with this car."

- **Emotional evaluation:** Such evaluations are from non-tangible and emotional responses to entities which go deep into people's state of mind. For example, the following sentences express emotional evaluations:

    "i love iPhone,"

    "i am so angry with their service people" and

    "this is the best car ever built."

Sometimes, sentiment ratings are referred to a combination of polarity and emotion:

Sentiment	Rational	Emotional
Positive	1	-2
Neutral	0	0
Negative	-1	2

Some Datasets & Resources

The "Tweet Emotion Intensity Dataset" data set taken from "8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA- 2017)". http://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html

|			
10000	How the fu*k! Who the heck! moved my fridge!... should I knock the landlord door. #angry #mad ##	anger	.938
10001	So my Indian Uber driver just called someone the N word. If I wasn't in a moving vehicle I'd have jumped out #disgusted	anger	.896
10002	@DPD_UK I asked for my parcel to be delivered to a pick up store not my address #fuming #poorcustomerservice	anger	.896
10003	so ef whichever butt wipe pulled the fire alarm in davis bc I was sound asleep #pissed #angry #upset #tired #sad #tired #hangry ######	anger	.896
10004	Don't join @BTCare they put the phone down on you, talk over you and are rude. Taking money out of my acc willynilly! #fuming	anger	.896

EmoBank: 10k sentences manually annotated with Valence, Arousal, and Dominance. Link: https://github.com/JULIELab/EmoBank

The emotional dimensions of **Valence**, **Arousal** and **Dominance** (VAD) are able to derive a **person's interest** (attraction), **level of activation** and perceived **level of control** for a particular situation from textual communication, such as emails.

Id	split	V	A	D	text
110CYL068_1036_1079	train	3.00	3.00	3.20	Remember what she said in my last letter? "
110CYL068_1079_1110	test	2.80	3.10	2.80	If I wasn't working here.
110CYL068_1127_1130	train	3.00	3.00	3.00	.."
110CYL068_1137_1188	train	3.44	3.00	3.22	Goodwill helps people get off of public assist...
110CYL068_1189_1328	train	3.55	3.27	3.46	Sherry learned through our Future Works class ...

Yelp Dataset

(https://www.yelp.com/dataset/challenge)
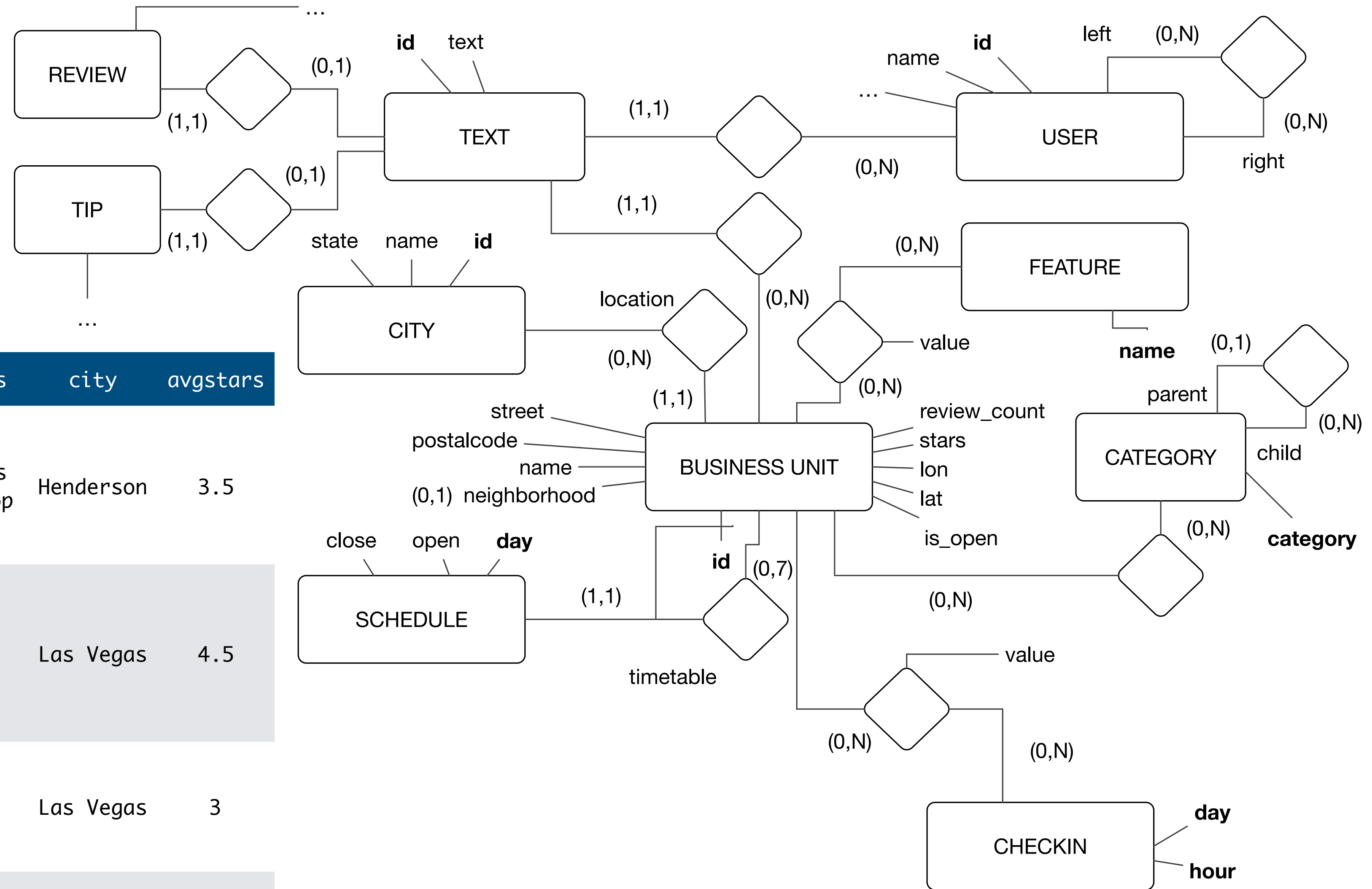
content	stars	business	city	avgstars
24 hour tacos? Yes please. This place is a modest spot in the middle of a strip mall, seats about 2 dozen, and has a little self serve …	3	Robertos Taco Shop	Henderson	3.5
2Nd time eating here today.1st time was great but now i dont think so. My husband ordered chicken fried steak. The gravey tasted very strange. …	1	Egg & I	Las Vegas	4.5
A word of caution! Don't use their ATM if you have a credit union. My Fiancé and I visited last Saturday. We expected to grab a to-go burger, just to say we …	3	Heart Attack Grill	Las Vegas	3
Absolutely amazing!! We just ordered take out and it was delicious (+was very well packed)...cannot imagine what its like to dine in :)	5	Khao San Road	Toronto	4

Multimodal Datasets

CMU-MOSI Dataset

The Multimodal Corpus of Sentiment Intensity (CMU-MOSI) dataset is a collection of 2199 opinion video clips. Each opinion video is annotated with sentiment in the range [-3,3]. The dataset is rigorously annotated with labels for subjectivity, sentiment intensity, per-frame and per-opinion annotated visual features, and per-milliseconds annotated audio features.

http://multicomp.cs.cmu.edu/resources/cmu-mosi-dataset

IEMOCAP Database

The Interactive Emotional Dyadic Motion Capture (IEMOCAP) database is an acted, multimodal and multispeaker database, recently collected at SAIL lab at USC. It contains approximately 12 hours of audiovisual data, including video, speech, motion capture of face, text transcriptions. It consists of dyadic sessions where actors perform improvisations or scripted scenarios, specifically selected to elicit emotional expressions.

https://sail.usc.edu/iemocap

Other datasets

Dataset	Author	Year	Description	Format	Size	Emotion categories
affectivetext	Strapparava & Mihalcea	2007	Classification of emotions in news headlines	SGML/txt	250 headlines	anger, disgust, fear, joy, sadnees, surprise, V
crowdflower_data	CrowdFlower	2016	Annotated dataset of tweets via crowdsourcing.	csv	40k tweets	anger, enthusiasm, fun, happiness, hate, neutral, sadness, surprise, worry, love, boredom, worry, relief, empty
dailydialog	Li Yanrand et al.	2017	Manually labelled conversations dataset with topics and emotions	text	13k dialogs	anger, disgust, fear, joy, sadness, surprise
emotion-cause	Diman Ghazi&Diana Inkpen&Stan Szpakowicz	2015	Automatically built dataset annotated with emotion and the stimulus using FrameNet's emotionsdirected frame	XML	820 sents + 1594 sents	anger, sad, happy, surprise, fear, disgust
EmoBank	Sven Buechel	2017	Large-scale corpus annotated with emotion according to VAD scheme	text	10k	VAD
fb-valence-arousal-anon	Preotiuc Pietro	2016	VA on Facebook	csv	2.8k posts	VA
grounded_emotions	Liu, V.&Banea, C.&Mihalcea	2017	They look into wheter the effect of weather, news events, relates to the tweet sentiment	text	2.5k tweets	joy, sadness
isear	Klaus R. Scherer and Harald Wallbott	1990	reported situations in which emotions were experienced	text (mdb/sav)	3000 docs	joy, fear, anger, sadness, disgust,shame, guilt
tales-emotions	Cecilia Ovesdotter Alm	2005	Dataset of manually annotated tales used in a document classification task	text	15k sents	angry, disgusted, fearful, happy, sad, surprised, mood (positive, negative)