

Guidelines for Annotation of Sentiment Corpus

Wladimir Sidorenko

October 20, 2013

1 Introduction

Welcome. In this assignment your task is to annotate sentiments in a corpus of Twitter messages.

1.1 Annotation Tool

For the annotation you will use MMAX2 - a freely available annotation tool that you can download under following link:

http://sourceforge.net/projects/mmax2/files/mmax2/mmax2_1.13.003/MMAX2_1.13.003b.zip/download

After you have downloaded and unpacked the archive file a directory called 1.13.003 will appear in your download folder. If you have never used MMAX2 before, please read chapters 1 to 5 of the document mmax2quickstart.pdf that you can find in the subdirectory 1.13.003/Docs. After reading this document, please change to your terminal in the subfolder 1.13.003/MMAX2 of the unpacked archive and type the commands

```
chmod u+x ./mmax2.sh
```

```
nohup ./mmax2.sh &
```

to start MMAX2. Then the MMAX2 window should appear on your desktop.

1.2 Corpus Files

You should also have received a copy of MMAX2 project files with the corpus data as a tar-gzipped archive. Unpack this archive using the command

```
tar -xzf twitter-sentiment.tgz
```

After that a directory called mmax-prj should appear in your current folder. Then go to your MMAX2 window and click on the menu File -> Load. In the popup menu select a path¹ to the unpacked mmax-prj folder and choose one of the *.mmax files there, say general.1.mmax. Now the source data will be loaded into MMAX2.

2 Tags and Attributes

In this annotation we will use the following tags with their according attributes.

2.1 <sentiment>

<sentiment>s are our most important markables. For the purpose of this assignment sentiments are statements that express a *polar attitude* of a writer or speaker to some objects, activities or events. We NEITHER consider vague emotional states of an author not related to any particular subject as sentiments NOR will we regard expressions of uncertainty or communicative acts like apologies, insults or commands as sentiments.

As a rule of thumb for annotating sentiments you should always act as follows: First ask yourself

¹Be sure the path itself contains no white spaces. Otherwise MMAX2 will fail to load the project file with the error message `java.net.MalformedURLException.no protocol: words.dtd`

whether a sentence you are looking at is neutral or expresses some polar attitude. In case it has some polar attitude ask yourself whether it is *distinctly targeted* at some thing or event that is visible or assumable from the context and if so mark the relevant phrase(s)/clause(s)² with the sentiment tag.

<sentiment>s also include phrases or clauses that express a polar opinion about some distinctly visible or at least assumable *targets*. These phrases/clauses should only contain information concerning the attitude and *whenever possible*³ not include any independent irrelevant information. Targets are usually expressed as:

- single noun phrases (e.g. *eine herrliche Blume* - *a wonderful flower*);
- single clause (e.g. *Ich mag diese Fernsehshow* - *I like this TV show*);
- multiple clauses (e.g. *Diese Regierung ist die beste aller Zeiten. April, April!* - *This government is the best ever. April fool!*).

2.1.1 Examples

Things you SHOULD mark as sentiments:

- Noun phrases and statements expressing polar opinions (e.g. *Der neue Film über Superman war knorke!* - *The new superman movie was fantastic*);
- Interrogative clauses containing a polar opinion and don't call this opinion into question but rather ask for its reasons or other related aspects (e.g. *Warum findet die Mehrheit der Bevölkerung CDU so toll?* - *Why does the majority of the population consider CDU great?*);
- Exclamations expressing support or disapproval of something (e.g. *Der Sommer ist da. Super!* - *Summer has come. Great!*).
- Comparisons⁴ expressing some preference of one object with regard to another (e.g. *Canon EOS 550d macht bessere Aufnahmen als 600d* - *Canon EOS 550d takes better pictures than 600d*).

2.1.2 Counter-Examples

Things you SHOULD NOT mark as sentiments:

- Statements describing some emotional states for which no target can be derived or found (e.g. *Ich fühle mich so traurig :(* - *Feeling so blue :(*);
- Statements describing some objective facts, even if possible consequences of these facts can be assumed to have negative influence on the author (e.g. *Wenn ein Floh einen Menschen beißt und ihn mit erbrochenem Blut infiziert, werden die Pestbakterien ins Gewebe übertragen.* - *When a flea bites a human and contaminates the wound with regurgitated blood, the plague carrying bacteria are passed into the tissue.*);
- Interrogative clauses in case they ask whether some polar opinion is true or not. (e.g. *Findet die Mehrheit der Bevölkerung CDU toll?* - *Does the majority of population consider CDU great?*);

2.1.3 Attributes

<sentiment> tags have following attributes with following values:

²Ich finde words, phrase oder clause passen besser als span

³Wie meinst Du das? Es wären also auch targets möglich, die irrelevante Infos beinhalten?

⁴Comparison sentiments are marked specially – see attribute *polarity* of *sentiment* tags and attributes *preferred/dispreferred* of markables *target*.

Attribute	Value	Value's Meaning
polarity	<i>positive</i>	this sentiment expresses positive attitude to its respective target
	<i>negative (default)</i>	this sentiment expresses negative attitude to its respective target
	<i>comparison</i>	this sentiment expresses comparison of two things with preference given to some of them
sarcasm	<i>true</i>	this polar attitude is meant as a derision, i.e. its actual polarity is the opposite of its apparent form (that means that an apparent praise is in fact meant as rebuke and vice versa - but this actual sense can only be derived on the basis of outside context or)
	<i>false (default)</i>	no sarcasm present - polar attitude has literal meaning
intensity	<i>0</i>	stylistically weakly marked expression which tends more towards a neutral sentence
	<i>1 (default)</i>	middle stylistic expressivity
	<i>2</i>	stylistically very expressive sentiment
sentiment-ref	<i>- > (directed edge)</i>	in case a single sentiment is not a contiguous text span but rather split into parts, you can draw an edge with this label from additional parts of a sentiment to its main part

2.2 <source>

The <source> tag is used to mark the original author(s) of the respective sentiment opinions. For example, in the sentence *Sie mag ihr neues Outfit nicht* (*She doesn't like her new outfit*) - "Sie" ("She") should be considered as opinion's origin i.e. the *source* of the sentiment relation.

Source tags should contain words/ phrases/ clauses denoting the author and are usually expressed by pronouns, nouns or multiword expressions, e.g. *Neue Züricher Zeitung*. Please mark complete noun phrases denoting the author(s), i.e. nouns with all their respective dependent attributes.

2.2.1 Examples

Things you SHOULD mark as sources:

- Original author(s) of polar opinions, i.e. those persons, groups or officials who experience sympathy or aversion for something e.g. *Michael meinte, das wäre die beste Lösung.* (*Michael thought, it would be the best solution.*);

2.2.2 Counter-Examples

Things you SHOULD NOT mark as sources:

- Persons how are neutrally citing someone else's opinion, e.g. in the sentence *Nach Tatjanas Worten war Michael sehr darüber verärgert.* (*According to Tatjana, Michael was very angry about that.*) Here you should only mark *Michael* as sentiment's source and not *Tatjana*. Please note, that in case an author supports or contradicts someone else's opinion, two sentiment relations should be made - one with the original author as source, and one with the supporter/opponent of the opinion enclosed in source tags.⁵

2.2.3 Attributes

source tags have only two attributes that both are directed edges. The first attribute called *sentiment-ref* should point from *source* elements to the *sentiment* span that this source belongs to in cases when: 1. multiple sentiment relations overlap each other, 2. the source element is outside the sentiment relation but is inside the visible context. The second attribute - *anaphref* - should only be applied in cases

⁵In that case sentiments will overlap, and you also should mark these sources with *sentiment-ref*-tag.

when the source element is expressed by a pronoun or a pronominal adverb and antecedents of them are visible within the local context. In that case, *anaphref* edge should point from pronoun/pronominal adverb to the respective antecedent.

Attribute	Value	Value's Meaning
sentiment-ref	– > (<i>directed edge</i>)	directed edge pointing from an outside source element to its respective sentiment, or in case of multiple overlapping sentiments when one source element is included in both of sentiment element - in that case this edge should point from source to its respective sentiment span
anaphref	– > (<i>directed edge</i>)	directed edge pointing from source element expressed by pronoun or pronominal adverb to their respective antecedents

2.3 <target>

As <target> you should mark persons, things, events or actions the opinion is made about.

Coordinatively connected noun phrases both of which represent targets should be marked separately. In cases the sentiment is a comparison, you should also mark two separate targets. One target should comprise the object that is preferred in comparison and the other one should encompass the dispreferred object. You should also set corresponding *preference* attributes in the respective <target> tags.

Targets are usually expressed by noun phrases or clauses. You are asked to mark complete noun phrases (that means nouns and all their respective attributes). Clauses should also be marked completely.

2.3.1 Examples

Elements you SHOULD mark as targets:

- Persons, things or objects the opinion is made about, e.g. *Ich hasse **lange Schlangen*** (*I hate **long queues***);
- Descriptions of actions or events, e.g. *Ich hasse es, **wenn Stärkere Schwächere drangsaliieren*** (*I hate it **when stronger people bully weaker ones***);
- Compared elements of a comparison, e.g. ***Ältere Handys** halten länger als **moderne Smartphones*** (***Older mobile phones** outwear **modern smartphones***).

2.3.2 Attributes

target attributes fully correspond to the attributes of *source* elements. So please refer to section ?? for the description of possible attributes for targets.

2.4 <emo-expression>

<emo-expression>s are lexical elements that bear some polarity meaning on their own. Usually this are words like *gut* (*good*), *schlechter* (*worse*), *mögen* (*like*), or *hassen* (*hate*). Such elements can be expressed either as single words or as complex expressions (in case of idiomatic expressions or support verbs).

2.4.1 Examples

Elements you SHOULD mark as emo-expressions:

- Adjectives and adverbs bearing polar attitudes *Peter hatte **bessere** Noten in der Schule als sein Bruder* (*Peter had **better** grades at school than his brother.*);
- Verbs expressing attitude of a speaker to target, e.g. *Mir **gefiel** die neue House-Staffel* (*I **liked** the new House series*);

- Idiomatic expression including support verbs e.g. *Zum Teufel soll die neue Regierung gehen* (*The new government should go to hell*).
- Smileys in case they really express an emotional attitude and are not used for politeness or without any particular meaning e.g. *Gleich in Braunschweig mit Kameraden treffen :) (Will soon meet friends in Braunschweig :))*.

2.4.2 Attributes

<emo-expression> tags have the following attributes with their corresponding values:

Attribute	Value	Value's Meaning
polarity	<i>positive</i>	this emotional expression has positive polar meaning about sentiment's target
	<i>negative (default)</i>	this emotional expression has negative polar meaning
intensity	<i>0</i>	stylistically weakly marked emotional expression
	<i>1 (default)</i>	middle stylistic expressivity
	<i>2</i>	stylistically very expressive emotional expression
emo-expression-ref	<i>– >(directed edge)</i>	in cases when a single emotional expression is not represented as a contiguous span but rather split into separate parts, this edge should point from parts to the main part of emo-expression (choose whatever part you want as main and connect all the other parts to it - it does not matter which part you consider as main)
sentiment-ref	<i>– >(directed edge)</i>	in cases multiple sentiments overlap each other and the emo-expression element is located in an overlapping span of both of them this edge should point to sentiment on which this emo-expression has immediate impact

2.5 <negation>

<negation>s are lexical or syntactic elements that reverse the primary polar meaning of emo-expressions to the opposite, so that the overall polarity of the whole sentiment is different to the polarity of emo-expressions belonging to it. A typical example of negation is *nicht* in sentences like *Ein guter Schritt war diese Entscheidung nicht*. (*This decision was not a good move*).

2.5.1 Examples

You SHOULD only mark negative elements that have impact on sentiment's polarity. Negating elements having no such impact should not be marked. Negation elements are usually represented by:

- Negation particle *nicht* (*not*), e.g. *Klug ist dieser Hund sicherlich nicht* (*This dog is certainly not clever*);
- Negative article *kein* e.g. *Er war kein Vorbild für seine Kinder* (*He was not a good role model for his children*);
- Indefinite pronouns like *niemand*, *keiner* etc., e.g. *Niemand hielt ihn für einen ehrlichen Menschen*. (*Nobody considered him an honest man*);
- Any lexical or idiomatic unit in case they turn sentiment's polarity to the opposite, e.g. *Ich zweifle, dass das neue iPhone ein besseres Display hat*. (*I doubt the new iPhone has a better display*).

2.5.2 Counter-Examples

DO NOT mark as negations elements that have no effect on sentiment's polarity. For example, in the sentence *Ich mag Leute, die nicht nur an sich selbst denken.* (*I like people who not only care about themselves.*) *nicht* (*not*) should not be marked as negation, since it does not change the positive polarity expressed by *mögen* (*like*).

2.5.3 Attributes

Negations only have one possible attribute, namely *sentiment-ref* that is a directed edge pointing from negation to sentiment it belongs to. You should only draw this edge in cases multiple sentiment relations overlap each other and it's not obvious to which of these sentiments the given negation belongs to, i.e. in cases the given negation is included in both sentiment spans.

2.6 <intensifier>

<intensifier>s are elements that increase the polar meaning of emotional expressions. Intensifiers are usually expressed by adjectives or adverbs like *sehr* (*very*), *ziemlich* (*rather*) etc.

2.6.1 Attributes

Intensifiers have following attributes and values:

degree	1 (<i>default</i>)	degree by which this intensifier increases polar sense of respective emotional expression (1 means slight increase), e.g. <i>ziemlich</i> , <i>recht</i> etc.
	2	this intensifier strongly increases polar sense of emotional expression, e.g. <i>sehr</i> , <i>super</i> , <i>stark</i> etc.
sentiment-ref	– > (<i>directed edge</i>)	a directed edge pointing from the intensifier to the sentiment it belongs to. By analogy to negations, you should only draw this edge in cases multiple sentiment relations overlap each other and it's not obvious to which of these sentiments the given intensifier belongs to, i.e. in cases when the given intensifier is included in both sentiment spans

2.7 <diminisher>

<diminisher>s are elements that decrease the polar meaning of emotional expressions. Like intensifiers diminishers are usually expressed by adjectives or adverbs. Typical examples of such adverbs are *wenig*, *kaum*, *ein bisschen* etc.

2.7.1 Attributes

Diminishers have the same attributes as intensifiers with the only difference in values of *degree* attribute. So here is once again a summary table of diminisher's attributes:

sentiment-ref	– > (<i>directed edge</i>)	a directed edge pointing from diminisher to sentiment it pertains to. You should only draw this edge in cases when multiple sentiment relations overlap with each other and it's not obvious to which of these sentiments given diminisher belongs to, i.e. in cases when given diminisher is included in both sentiment spans
degree	-1 (<i>default</i>)	degree by which this diminisher decreases polar sense of respective emotional expression (1 means slight decrease), e.g. <i>wenig</i> , <i>bisschen</i> etc.
	-2	this diminisher strongly decreases polar sense of emotional expression almost turning the sense of emo-expression to the opposite, e.g. <i>kaum</i> etc.