Evaluative and Probabilistic Language Annotation Guideline

June 28, 2024

1 Annotation Description

The annotation task comprises several components that need to be adressed. The main target of annotation is lexical items that bear evaluative meaning or a positive or negative connotation, such as *struggle* or *surprisingly*.

A tag for the presence of tokens bearing evaluative meaning (<EVAL>) is needed, as well as a set of tags that represent the part of speech of the respective evaluative expression, such as VERB, ADJECTIVE or ADVERB. I propose to use the Stuttgart-Tübingen Tagset or the Penn-Treebank tagset for this purpose.

According to the work presented in (Langner & Klabunde, 2024), further elements that need to be annotated are the polarity of the evaluative expression {POSITIVE, NEGATIVE, NEUTRAL}, the evaluated factoid or claim (corresponding to an information piece in the database), which we call the target feature, optionally any features the authors name as the background of their evalaution, which we call the source feature, and finally the reference of the evaluation.

category	tag	values
POS-tag	<POS>	see Stuttgart-Tübingen /Penn-Treebank
evaluative item	<EVAL>	empty
probability item	<PROB>	below_threshold, above_threshold
polarity	<POL>	{POSITIVE, NEGATIVE, NEUTRAL}
eval. target feature	<E-TARGET>	empty
eval. source feature	<E-SOURCE>	empty
prob. target feature	<P-TARGET>	empty
prob. source feature	<P-SOURCE>	empty
context	<context>	extendible
reference_type	<ref_type>	{ STARTER, BENCH}

Levels of annotation:

TAG	Tagset	layer type	layer spec. in inception
POS	Penntree	span annotation	(Universal Pos tag pre-defined)
EVAL_ITEM	None	span annotation	type=span, granularity=token-level. overlap=any
PROB_ITEM	None	span annotation	type=span, granularity=token-level. overlap=any
POLARITY	{POS, NEG, NEUT}	span annotation	type=span, granularity=token-level, overlap=any
E-TARGET	db features	span annotation	type=span, granularity=token-level. overlap=any
E-SOURCE	db features	span annotation	type=span, granularity=token-level. overlap=any
P-TARGET	db features	span annotation	type=span, granularity=token-level. overlap=any
P-SOURCE	db features	span annotation	type=span, granularity=token-level. overlap=any
REFERENCE	None	span annotation	type=span, granularity=token-level. overlap=any
COMMENT	free-text	span annotation	type=span, granularity=token-level. overlap=any
context	extendible	span annotation	type=span, granularity=token-level. overlap=any
reference type	bench or player	span annotation	type=span, granularity=token-level. overlap=any

Table 1: Caption

EXAMPLE SENTENCE:

[Surprisingly]$_{EVAL\_ITEM,pos,ADV}$, [player A]$_{REF}$, despite his short time of [15 minutes]$_{E-SOURCE}$ on the field, scored [impressive]$_{EVAL\_ITEM,pos,ADJ}$ [35 points]$_{E-TARGET}$ in this game], against player B's 22 points.

Layer explanations:

- EVAL_ITEM: this is a binary tag that can span one or more tokens. It is used for marking words with evaluative meaning.
- PROB_ITEM: this is a binary tag that can span one or more tokens. It is used for marking words with probabilistic meaning.
- POS: The POS level provides part of speech tagging for evaluative and probabilistic items. This means, that any token sequence annotated with EVAL_ITEM or PROB_ITEM needs POS tagging, but no other elements.
- POLARITY: the polarity tag informs about the polar stance of the evaluative items. Polarity may be positive (impressively), negative (disappointingly) or neutral, which is also a bin for unclear cases which are neither clearly positive or negative (e.g. unsurprisingly).
- E-TARGET: this layer marks the feature from the database, e.g. the points score, rebounds, etc., which is being evaluated, e.g. in the example sentence, the 99 points score of team A is evaluated as impressive, therefore "99 points" are the target feature.
- E-SOURCE: The source feature is the feature on basis of which an evaluation is formulated. Source features are optional, they are not always explicitly named. In the example sentence, "the short time of 15 Minutes on the field" is the source feature, in regard to which 33 points are impressive.
- P-TARGET: this layer marks the feature from the database, e.g. the points score, rebounds, etc., which is being probabilistically assessed, e.g. in the example sentence, "surprisingly" bears probabilistic meaning, "99 points" is again the target feature.
- P-SOURCE: The source feature is the feature on basis of which a probabilistic assessment is formulated. Source features are optional, they are not always explicitly named. In the example sentence, "the short time of 15 Minutes on the field" is the source feature, in regard to which 33 points are "surprising".
- REFERENCE: The reference of an evaluative (or probabilistic) item is the named entity (players, teams etc.) to which the target feature is related, e.g. in the example sentence, "player A".
- LINKAGE: The linkage layer is an auxiliary layer which is supposed to connect E-TARGET, E-SOURCE, REFERENCE and POLARITY with the central element EVAL_ITEM. So each of the elements listed above has a linkage tag, these linkage tags are then connected to form a chain (see inception screenshot below).
- reference_type: one very important element of evaluative expressions in relation to single players is whether they are starter players or a reserve player from the bench. This is annotated separately in case the reference refers to a single player because of its extraordinary importance for analysing a large number of cases.

- context: the context gives further information on certain evaluative expressions in regard to the choice of the contrast set, mostly in regard of the time frame certain items refer to, e.g. the performance of a team in the first three quarters. The context is related to SOURCE, but differs in such that the context refers to those where the evaluative background is numerically not specifiyable, but has to be mapped on a set of categories from prose descriptionn in the text.

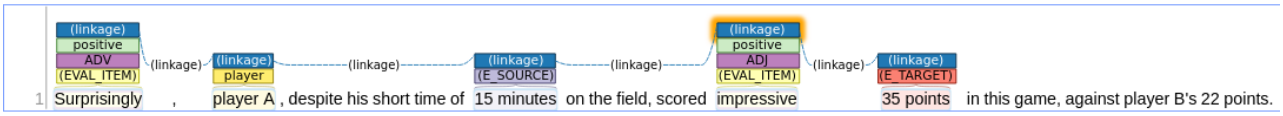

Figure 1: Caption

Known Issues

contrasts In the given example, there is a contrast relation to the points score of some other player B (against player B's 22 points). In case there is no connection to the annotated evaluative context, it is better to leave them unannotated

items when several evaluative or probabilistic items are contained in a sentence, they are distinguished on the basis of what they modify. If an item modifies another evaluative item, e.g. "assiduously well-balanced", then they are judged as one item, while having two different POS tags. IF they are split as "surprisingly" and "impressive" in the example sentence, they should be annotated as two different items that share the same target feature, reference etc., just in the way the screenshot depicts.

missing elements when an element is not present, e.g. the source feature or sometimes even the target feature, when the evaluative item describes the global performance of a player, only the present elements are annotated, even though this produces incomplete or underspecified items.