

“Marketing is about values”





From unclean Facebook Comments to Clean Data for Profiling

Sandro Skansi, IN2Data

Argumentation Mining

- One of the newest areas of Natural Language Processing
- Basic idea:
 - Get the input text and knowledge base
 - Take out the hidden premises

Argumentation Mining

- Example:
 - From: **Speaker S**: “I don’t like Obama, since he doesn’t support a small government.”
 - Get: **S**: `For(Res_Fisc_Pol)`
 - From: **Speaker T**: “Economy grows, don’t let republicans fool you”
 - Get: **T**: `For(Exp_Fisc_Pol)`
- This is currently under active development (EMNLP2016: Boltuzić, Skansi, Šnajder)

Knowledge Base

- The system depends on the Knowledge Base
- This base contains the ontology and common premises:

Government \subseteq Economics

SmallGov \Rightarrow Res_Fisc_Pol

EconGrowth \Rightarrow Exp_Fisc_Pol

Against_S(P) & Against_S(Utterance_P) \Rightarrow For_S(Utterance) [NXOR]

Exp_Fisc_Pol & Res_Fisc_Pol \Rightarrow False/Conflict [field of conflict: Economics]

- We derive JSONs:
 - {name: S, mostConflictsIn: Economics, positions: [For(A), Against(B),...]}
 - {name: T, ...}

Procedure

- Start: S: “I don’t like Obama, since he doesn’t support a small government.”
- Parsing: [don’t, like, Obama] [doesn’t, he, support, small government]
- NER: Obama he, small government
- SA: Obama, -1 he, SmallGov, -1
- AnaphRes: he→Obama
- RelExtr: Against(Obama) Against(Obama,SmallGov)
- Tidy-up: $\text{Against}_s(\text{Obama})$ $\text{Against}_s(\text{SmallGov}_{\text{Obama}})$
- Our module: SDL-Res and ILP-based (W, V)
- Result: JSON

Difficulties

- The system relies heavily on the knowledge base
- The size is not as much as problem as one might think, since the exponential growth is there only at the early stages.
- In principle, the base should not be hardcoded but learned by some clustering method over data:
 - This is an open problem: so far, the closest we have are IR-webcrawlers
- How powerful the background logic: clausal, propositional, FOL, HOL? Fragments?
- In our example we went forward, but there will be a need for ILP not just regular resolution.

Values

- For every speaker we have a list in JSON:
 - For(A)
 - Against(C)
 - For(B)
 - ...
- Based on this, we can conduct clustering/multinomial classification to put the person in a box, based on his comments.
- This sounds harsh, but the final purpose of this is giving a personalized offer based on the person's values
 - e.g. eco-friendly products to a green person

THANK YOU FOR YOUR ATTENTION

QUESTIONS ARE WELCOME

NOW, OR VIA EMAIL

sandro.skansi@in2data.eu