Semantic rules based

Preclude approach: extract clause action/temporal/effect/conditional clauses/phrases using lexicons of potential indicators (my take away from this is their parser only works on their corpus:

Semantic patterns, key preposition indicates location or temporal clauses

It should be noted that the semantic rules described in sections IV-A and IV-B are **novel, customized to health advice statements, and guided by linguistic patterns**. They enable Preclude to extract the semantics of advice statements and accurately detect the conflicts as demonstrated in section V.

After assigning polarity to the semantic tokens of an advice, the problem is reduced to mapping the token sets to the potential cases of contradiction presented in Section II.

We are not detecting the conflicts from case 7 in this work, as it requires additional physiological data of a user and accurate modeling of the effects of different health interventions that are currently unavailable.

For labeling ground truth, objects are manually extracted from each advice by 3 human annotators.

Semantic interpretation with external knowledge base

Manually learn and tune model? Ask Sarah about Preclude (seems like manually tuned model)

*NLP technologies:*

WordNet (Verb synset, Noun synset, hypernym and hyponym), SUTime, POS tagging, Dependency Parsing

*Priority*:

1. Build a corpus

Reason: Non-trivial and gathered from professional sources. We can make it extreme trivial and simple to a point of direct mapping with rudimentary parsing techniques. Corpus also has to be credible like the one in Preclude paper.

1. From Preclude paper, narrow objective (domain specific lexicons): 1156 health advice only, and build lexicons base on their own collected data, may not be generalized to other area. They also found out general-purpose parser degraded performance (phase 3 c), customized corpus -> customized parser

Essentially not translating any given English sentence but domain-specific: “I love vanilla and my mom makes awesome pumpkin pie” won’t show up in the corpus

1. Can be regarded as a contribution: first of its kind: smart city service specifications
2. Easy to make empirical observation and analysis
3. Make the scope of study clear
4. Easier to argue validity and contribution: why care

How-to:

1. Paraphrase existing sentences to express the same STL semantic meaning

(e.g. “The average waiting time of vehicles in a lane should not increase by more than 10%” vs. “No vehicle should wait for more than 10% of…

1. Given an STL generate several English specification
2. Compose new sentences with given sensor, duration, threshold, location, and etc.

Cases to consider: “The noise level in a lane should always be less than 70 dB” and “the decibels in a lane should be less than 70 dB”

1. Labeling ground truth: manually convert to STL
2. Once finished building parser and ran test, do analysis by parts (threshold info extraction accuracy and etc.) and by whole (overall conversion rate)

*Challenges:*

1. Semantic challenge: Need to extract implied sensor/signal. “No collision should happen” Vs. “No vehicles should collide” both convert to “Collision” signal/sensor.
2. Semantic challenge: Need to extract implied relation. “The highway should be unblocked within 30 minutes” Vs. “The highway should not be blocked by more than 30 minutes.” Vs. “The highway should be blocked by less than 30 minutes”
3. Conceptual overlap: Emergency Vehicle Vs. Ambulance, Police car, etc…
4. Conditional? “If it’s raining, turn on the street light and set the illumination to level 3.”
5. Missing parts: no duration, no threshold
6. Compound nouns or verbs (vehicle collision, turn on)

*Problem formulation*: (What are we looking for in a sentence?)

Sensor/Signal:

Vehicle Number, Congestion, Vehicle Wait Time, Pedestrian Number, etc…

Actuator/Action??

Direct, turn on, set, etc…

Relation:

<, >, ≤, ≥

Space/location:

1. General Area: school zone, residential area, intersection, street, lane, block, etc…
2. Specific Location: on Emmet Street, US-250, around UVA campus, etc…

Duration:SUTime package

1. General time frame: no more than 30 min, less than 5 seconds, within 10 min,
2. Specific time frame: 5 AM to 2 PM.
3. Implied time frame: after some point, after midnight,

Threshold

1. Temporal threshold: ≈ General time frame duration,
2. Quantity threshold: 10 mg, 70 dB, etc…

*Solution:*

Convert a specification to a tuple: Specification: <Sensor, Relation, Space/Location, Duration, Threshold>

1. Temporal Clause: Prepositional Phrases (PP)

Q:

1. R9: blocked(highway) < 30 since no Until statement?
2. Unit issue do we retain unit or convert to a standard?
3. Historical values/constants
4. Traffic congestion in a lane should not increase by more than 20%. Why is it compared against its historic value not current value? Or why not substitute Yield’(lane) with a numerical value.
5. Is it possible to be conditional? How does STL handle conditional statement? How does sumo handle conditional statement?
6. We need corpus, a rich sample of possible specification statements, to make observation and empirical conclusion
7. Preclude uses existing corpus: online dataset and apps
8. Ground Truth labeling
9. Manually learn and tune model? Ask Sarah about Preclude (seems like manually tuned model)