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, Named Entity Recognition, Stanford Core NLP (Java NLP library)

*Priority*:

1. Build a corpus

In Preclude, they have: “a health advice dataset comprising of 1156 health advice statements covering 8 important health topics.” We should something similar. A number of smart city service specifications covering x number of domains.

Basic assumptions:

Non-trivial and gathered from credible sources. Complete and grammatically correct English sentences. For preliminary Study no multiple sentences for one specification.

Reasons:

1. From Preclude paper, narrow down objective (build domain specific lexicons and parsing components): 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 requirements: “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. Help to define the problem and make the scope of study clear
4. Easier to argue validity and contribution: why care

How-to:

1. Refine signals/sensors definitions:
   1. Proper names: R5 WaitTime -> VehicleWaitTime, R6 Pedestrian -> PedestrianNumber
   2. Parameters: Space/Location, general/specific
   3. Return type: seconds, minutes, mg, integer, etc…
   4. Domain: Transportation, Emergency, Environment…
2. 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.
3. Maybe crowdsourcing? English major students and/or Civil Engineering Students?

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:*

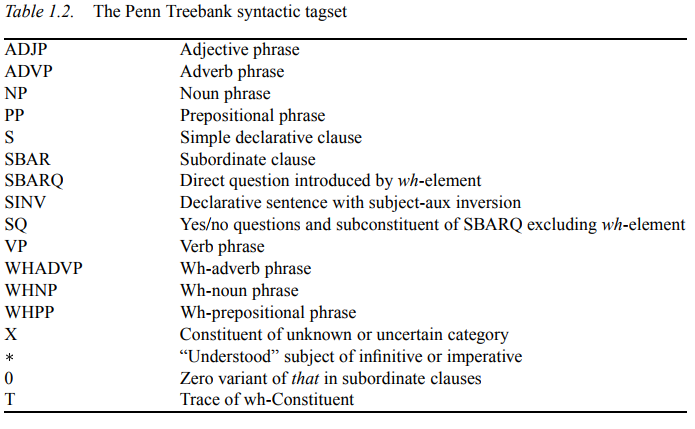
Convert a specification to a tuple: Specification: <Sensor, Relation, Space/Location, Duration, Threshold>

1. Temporal Clause: Prepositional Phrases (PP)

*Other points:*

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. Ground Truth labeling
7. Manually learn and tune model? Ask Sarah about Preclude (seems like manually tuned model)

*Observations or assumptions:*

1. Duration is more likely to be associated to a time threshold.
2. 1 – to – 1 mapping. One clause contains one sensor.
3. There are two types of sensors: one that cannot be expressed with a verb: vehicle number, and one contains or can be expressed with verb: vehicle wait time.
4. 
5. for the time being we don't consider compound-complex sentence.

Citations:

Stanford Core NLP website has citation information for different tools.

MIT Word Net API citation: <https://projects.csail.mit.edu/jwi/>

Table 1.2. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.9.8216&rep=rep1&type=pdf

*Pseudo Algorithm:*

Step 1: Determine Sentence Structure: Run through constituency parser

Simple Sentence

Compound Sentence (… and …), expressing two actions in one sentence

Conditional Sentence (if … then …)

Compound Complex

Step 2: break it down to clauses

If compound, determine the coordinator (FANBOYS). In the context of smart service specification, only consider “and”, “nor”, “but”, “or”, and “yet”

If complex, dependent clause dictates the pre-condition or →

Step 2: Named Entity Extraction

Step 3: Regular Expression pattern matching for sensors/actuators:

Reason: 1. formality variations: degree of ≈ level of

1. Semantics variations: vehicle ≈ car ≈ automobile (WordNet synsets manual search expansion)

Step 4: Relation extraction

Step 5: Location info extraction

Step 6: Time duration (Not the same as the NER duration)

Programming Techniques:

Singleton pattern: customized parser: ParserUtility Class NLP pipeline is very expensive, eliminate multiple instances, and most are utility functions