Print

Thứ ba, 18 Tháng sáu 2019, 8:29 PM

Hệ thống: **Trường ĐH CNTT - Website môn học**
Khoá học: **Xử lý văn bản Y khoa - CS339.J21.KHTN (CS339.J21.KHTN -- **Clinical NLP**)**
Bảng từ: **Health NLP Terms-Glossary for CS339**

A

Annotation

Annotation systems are the markup styles used by NLP systems (called machine annotation) and by humans (called human annotation) to highlight target NLP mentions or other features in a text, like the start/stop of a sentence, an NER term, or a part-of-speech. When an NLP system annotates text this is called labeling ("to label a mention").
Each NLP framework has its own annotation syntax. UIMA, CLAMP, and GATE all use a distinct form of annotation. Human annotation is usually considered the reference, or *gold*, standard that NLP tries to match. Comparing human annotations with your NLP system annotation is called "performance metrics". There are several performance metrics in NLP -- see the glossary entry.
Usually annotations are stored separately from the original document being analyzed. These are called "stand off" annotations. The most most common uses a character offset, where the first character of a document is "0", the next character is "1", etc. The annotation is a pair of numbers:the offset of the first character of the text being annotated and the offset of the last character, For example, the first word of **this** definition, "Annotation", would be 0-9. The "A" is at offset 0 and the "n" is at offset 9.

Assert

An instance of NLP system output, i.e., the NLP system finds what it is looking for and classifies it.
For example, in the sentence “John Hurdle has temperature of **40 C**.” the NLP system would assert an instance of **fever**. In this example, the NLP system has annotate "40 C", interpret it a temperature, and classify a temperature over 37 C as fever.

Assertion

The noun form of "assert". If an NLP system asserts fever in the sentence, "John had a temperature of 40C", then "fever" is the assertion.

C

CLAMP

Clinical Language Annotation, Modelling and Processing tool kit; an NLP pipeline tool set offered by the lab of Dr. Hua Xu of the University of Texas (Houston). We thank Dr. Xu and his team for their support. CLAMP requires only minimal programming to use. It is based on a drag-and-drop GUI.

Context

In health NLP, context is means whether a term mentioned in a clinical text is a term of interest to an NLP system. These are the most important context issues in health NLP:
Negation: if the term is excluded, e.g., "The patient does *not* have cancer."
Experiencer: if the term refers to the current patient or to someone else, e.g., "Her *father* had cancer."
Historical: if the term applies to the current note or to the past. "The patient *had* cancer three years ago but now is cancer *free*." The first instance of "cancer" is historical, the second instance of "cancer" is a negation example.
Probable: if the term is uncertain or not confirmed, e.g., "The patient *may have* cancer but we need more testing to be sure."

Corpora
The plural form of "corpus". (see Corpus)

Corpus
A collection of texts processed by an NLP system. Examples of a corpus in health NLP:
All of the clinical notes belonging to a patient (small corpus).
All of the notes belonging to an entire hospital (big corpus).
All articles ever written on natural language processing (very big corpus).

F

FN
False negative (see "Performance metrics")

FP
False positive (see "Performance metrics")

Framework
An NLP framework is specific approach to NLP processing. Typically, a framework is defined by its underlying tool set and its annotation system. UIMA, GATE, and CLAMP are examples of NLP frameworka.

G

GATE
GATE, Generalized Architecture for Text Engineering, is a Java-based open-source NLP framework and tool set created at the University of Sheffield. It is one of the most mature frameworks available. Its use is not limited to clinical applications, but it has no special tools for clinical NLP.

H

Healthcare system
All hospitals and clinics. The Vietnamese healthcare system is all hospitals and clinics in the country.

M

Machine learning
Machine learning, abbreviated ML, is programming computers so they can learn from data. Simple to define, hard to do!
Machine learning is used in two ways in health NLP:
Some NLP systems use ML to help find mentions in text. They might learn, for example, that every time the pattern *<number>"C"* occurs in a clinical note, the <number> means a temperature value. "John had a temperature of 40C and was sweating."
After finding mentions in text, ML is often used to make higher level assertions at the note, encounter, or patient level. For example, if "fever" is asserted multiple times in a patient's collection of notes, then the system might assert that the patient has "chronic fever." That is an example of a patient-level assertion.

Mention
An occurrence of term of interest in one or more texts being analyzed by an NLP system. In health NLP, a "mention" == an NLP "target". For example, assume this line appears in a document in a clinical corpus:
"The pt described a history of **ear infections** and had a **fever**."
This line has a *diagnosis* mention/target ("ear infections") and a *symptom* mention/target ("fever"). Note that the "ear infections" mention/target refers to diagnosis made in the past. This is an example of *mention context*.

N

NER
Stands for "named entity recognition." NER is an NLP task that finds mentions of specific terms of interest to a specific NLP project. Clinical NER includes medical terms not usually found in everyday texts. The most common types of medical terms are: **medication names** (like "penicillin" or "paracetamol"), **allergies** (like "reaction to penicillin" or "pollen allergy"), a **diagnosis** (like "heart attack" or "stomach cancer"), a **symptom** (like "fever" or "dizzy"), and **medical tests** (like "serum glucose" or "x-ray").

Clinical NER is complicated by the use of short forms like acronyms (like "MI" for "myocardial infarction" or "ECG" for "electrocardiogram") and abbreviations (like "glu" for glucose or "pt" for patient).

NLP
Natural Language Processing; the theory and practice of computer analysis of texts generated by humans.

Note
A description of a patient encounter written by a nurse, doctor, or other member of the healthcare system.

P

Parsing
Parsing is an NLP task that extracts part-of-speech, or chunks of speech, from texts. The VnCoreNLP system is an example of a Vietnamese parser. The Stanford Parser is an example of an English parser (one of the best).

Patient
A person cared for by the healthcare system.

Patient encounter
An episode of care, i.e., when a patient visits doctor, clinic, or hospital.

Performance metrics
These are standard ways to determine how well an NLP system is working. They are all based on comparing the NLP output of your system to a reference (or gold) standard. The reference standard is frequently built from human annotations.
For example, assume your NLP system is processing 1,000 texts. And assume that in those texts there are 100 clinical-term mentions of interest to your project (like 100 diagnoses and your system is trying to find *all* diagnosis mentions). To run a performance metric, a human has to **annotate these mentions.** These 100 mentions are called "true positives (**TP**)". Also assume that there are 200 *other* clinical terms that are not diagnoses. These are called "true negatives (**TN**)". When your system labels a TN as a positive, this is called a "false positive (FP)". When your system labels a TP as a negative (in this example, a non-diagnosis term is labeled as diagnosis), this is called a "false negative (FN)". Once you understand TP, TN, FP, and FN, you have everything you need to compute most important performance metrics.
The most popular metrics are:
- Accuracy: fraction of all the labels assigned by your system that are correct.(TP + TN)/(TP + TN + FP+ FN)
Accuracy can be useful, but if TP are rare, then simply labeling *every* mention as a negative mention gives a high accuracy.

- **Precision:** the fraction of all your system's positive labels that are *really* positive.
TP/(TP + FP)
Precision goes up as false positives go down. This is also called "positive predictive value."

- Recall: the fraction of all the positive mentions you *should* have found that were labeled as positive.
TP/(TP + FN)
Recall goes up when the false negatives go down. This is also called "sensitivity."

- F-score (or F-measure): a combined measure that balances precision and recall equally.
2*[Precision * Recall)/(Precision + Recall)]

Why is F-score a useful measure? The closer recall and precision are to each other, the more like a simple average this measure becomes. But the more recall and precision differ, the smaller the F-score. This penalizes a system that is good at precision *or* recall but not good at *both*. Technically, this is the "harmonic" mean of precision and recall. Note that *some* NLP applications *might* want to weigh recall higher than precision (where it is important not to miss positives) or weigh precision higher than recall (where it is expensive to label a positive in error).
There are many other metrics, and there are ways to adjust F-score to weigh recall and precision differently. For a good review, see this Wikipedia entru.

Pipeline
A pipeline is an approach to NLP that uses a series of processing modules to analyze texts. Each module performs one task, like sentence segmentation or POS parsing. Typically, a module takes input from the module before it and delivers output to the module after it. This is the most common framework used in clinical NLP today. The modular design allows for quick updates (i.e., to change a parser you need only change the parsing module). Some pipeline frameworks, like UIMA-AS, support parallel stages. Normally this is used to speed up a stage that is a performance bottleneck. CLAMP is the pipeline we use in this class.

POS
An acronym that stands for Part Of Speech. A part-of-speech is a grammatical concept of human speech. Human languages have a general syntax. That syntax is not as rigorous as a computer language. Parts-of-speech are connected by an underlying grammar unique to each language.Human languages share parts-of-speech like verbs (action words) and nouns (things). Typically, POS are extracted from text by a parsing the text.

S

Segmentation
To break a text into individual sentences (sentence segmentation) or sentences into words or phrases (phrase segmentation).

T

TN
True negative (see "Performance metrics")

TP
True positive (see "Performance metrics")

U

UIMA

Unstructured Information Management Architecture, a framework from IBM that can be used for NLP. UIMA is not limited to NLP, however. A key aspect of UIMA is its annotation type system. Each stage of a UMA pipeline performs one (typically) kind of annotation, like POS tagging, storing the annotations separately from the text. Input texts are unchanged as they are passed along the pipeline stages. IBM used a version of UIMA, UIMA-Asynchronous Scaleout, in its Watson system. Watson is used in health NLP, but many of IBM's claims for this system are not well supported.

V

VnCoreNLP

This is an NLP pipeline designed for the Vietnamese language. Inspired by Stanford's CoreNLP tools, VnCoreNLP offers word and sentence segmentation, POS tagging, and NER -- all designed specifically for Vietnamese texts.