Supplementary: Improved Pattern Learning for Bootstrapped Entity Extraction

1 Drug & Treatment Class Definition

In this paper, we evaluate learning drug and treatment (DT) entities from medical text. Similar to Anonymized (2014), a DT entity is defined as a pharmaceutical drug, or any treatment or intervention mentioned that may help a symptom or a condition. It includes surgeries, lifestyle changes, alternative treatments, home remedies, and components of daily care and management of a disease, but does not include diagnostic tests and devices. If an entity is a pharmaceutical drug, it is considered a DT, regardless of whether it helps or causes a particular symptom as a side effect. The class includes all relevant generic terms (for example, 'meds', 'over-the-counter') and devices to improve a symptom or condition (like 'inhalers').

2 Examples

Some example sentences from the dataset are following. The words in bold are considered as DT.

> I heard that the stem of certain **Opuntia** is used to treat type II diabetes, diarrhea, and stomach ache.

> Have you considered using a good quality **hepa filter**?

> Apply an **astringent toner** only to the oily areas and a light **moisturiser** only to the dry areas.

> The PNS and Adenoids xray showed minor opacities in the region.

> To solve the blocked e-tubes i used **cortison spray** and did the **valsalva maneuver** several times a day for at least two months.

3 Labeling Guidelines for Annotators

For class DT, annotators were asked to label entities belonging to DT as positive and all others as negative. They queried 'word + forum name' on Google and manually inspected the results. Apart from the definition of the DT class above, the following instructions were followed for each class.

3.1 Positive

The following types of variations of DT entities were allowed: spelling mistakes, abbreviations, and phonetically similar variations (for example, 'brufen' for 'ibuprofen'). If a word or phrase was a part of a DT entity, then it was labeled positive. For example, 'nux' is considered positive because 'nux vomica' is sometimes used medicinally. Generic entities that can be used as a treatment for the medical condition were included (like 'moisturizer' for Acne). Brand names of DT entities, like 'Amway', were labeled positive. Ways to administer a medicine were included, such as 'syrup', 'tabs', and 'inhalation'. Phrases like 'antibacterial' or 'asthma meds' were also considered positive.

3.2 Negative

Entities that were not labeled as positive were considered negative. If a phrase had any non-DT word then it was considered negative, except when phrases had the name of the disease or symptom for which the treatment is mentioned. For example, 'sinus meds' was considered positive. Websites, dosages, diagnosis tests or devices, doctors, specialists were also labeled as negative.