Project 2: Identifying Tweets with Adverse Drug Reactions

COMP90049
Knowledge Technologies

Jeremy Nicholson and Sarah Erfani

Semester 2, 2017

THE UNIVERSITY OF MELBOURNE

Adverse Drug Reactions

(Pharmacological) drug:

> *... a chemical substance used in the treatment, cure, prevention, or diagnosis of disease or used to otherwise enhance physical or mental well-being.*

(`http://www.dictionary.com/browse/drug`)

Adverse Drug Reactions

Adverse drug reaction:

> *... an injury caused by taking a medication.... The meaning of this expression differs from the meaning of "side effect", as this last expression might also imply that the effects can be beneficial.*

(`https://en.wikipedia.org/wiki/Adverse_drug_reaction`)

Tweets

Tweet about a drug:

> *This Vyvanse needs to kick in.*

(No ADR)

> *I have got to stop taking my Vyvanse so late!! #nosleep #addproblems*

(ADR - insomnia)

Project 2

Supervised Machine Learning:

Given some tweets identified as containing an ADR (`Y`) or not (`N`), can we build a system which can successfully predict whether a given (unseen) tweet contains an ADR?

Project 2

Basic idea: tokens in tweets are somewhat indicative. Therefore:

1. Build a VSM over the tweets
2. Use token frequencies as features
3. Train a model
4. Evaluate the model

Project 2

Problem 1: many terms (e.g. `in`) are not indicative

Problem 2: too many terms (about 9K)

Project 2

Basic idea: tokens in tweets are somewhat indicative. Therefore:

1. Perform Feature Selection, get best 92 terms (done)
2. Build a (reduced) VSM over the tweets (given in the ARFF files)
3. Use token frequencies as features (given in the ARFF files)
4. Train a model (trivial using Weka, or similar)
5. Evaluate the model (trivial using Weka, or similar)

Project 2

Basic idea: tokens in tweets are somewhat indicative. Therefore:

1. Perform Feature Selection, get best 92 terms (given in the ARFF files)
2. Build a (reduced) VSM over the tweets (given in the ARFF files)
3. Use token frequencies as features (given in the ARFF files)
4. **Find some other feature(s)**
5. Train a model (trivial using Weka, or similar)
6. Evaluate the model (trivial using Weka, or similar)
7. **Gain some knowledge!**