On the Ground Validation of Online Diagnosis with Twitter and Medical Records

Todd Bodnar[*]
Pennsylvania State University
Department of Biology
University Park, PA 16802
tjb5215@psu.edu

Maybe Conrad, Maybe Vicky

Marcel Salathé
Pennsylvania State University
Department of Biology
University Park, PA 16802
salathe@psu.edu

ABSTRACT
This is an abstract

Categories and Subject Descriptors
I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems—*Medicine and Science*

General Terms
Experimentation, Validation

Keywords
Twitter, Validation, Digital Epidemiology, Remote Diagnosis

1. INTRODUCTION
2. RELATED WORK
3. DATA COLLECTION
3.1 Medical Records
3.2 Twitter Records
4. SIGNAL DETECTION
4.1 Event Based Signals
Most work to this point considers finding messages in tweets (i.e. "I'm sick") or in keyword frequencies.

Consider keywords, we try using an unsupervised method to find words that seperate months with and without ILI.

Now we try using hand chosen keywords, see if we can distinguish between individuals' sick and healthy months.

Finally, we hand rate all tweets in sick months to see if we can find tweets announcing illness. Find % of individuals do tweet about being sick. Also test subsample of non-sick months to see for false positives.

4.2 Frequency Based Signals
Look at changes in behaviour based on illness. May have signal even if no relevant messaging.

In each user, take months when they did tweet, apply normalization (val - min)/(max-min)

Build distributions of months before, during, and after illness. Compare distributions. Try paired / unpaired

Fail to find sig difference between three sets → comment on benefits of this.

Try anomoly detection...

4.3 Network Based Signals
Preliminary idea. Cascade effects causing echoes on social network. Also consider friends becoming ill around same time. Check @ tag

5. ANALYSIS
6. CONCLUSIONS
7. REFERENCES

[1] L. R. Ford and D. R. Fulkerson. Maximal Flow through a Network. *Canadian Journal of Mathematics*, 8(3):399-404, 1956.

8. REFERENCES

---

[*]Corresponding author