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

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.
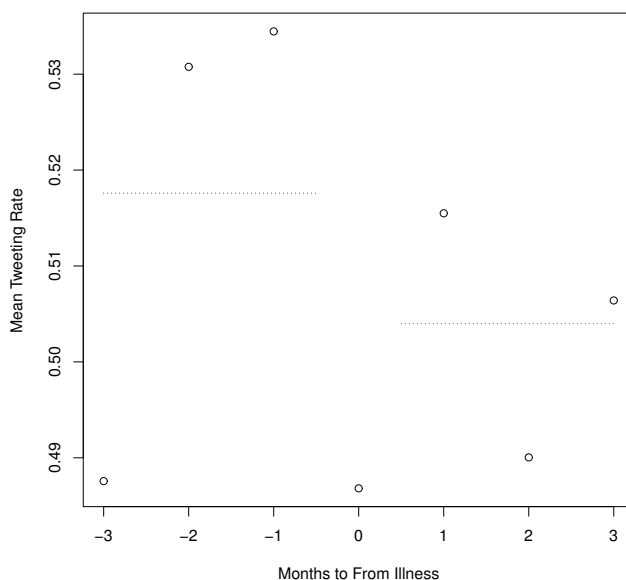
---

[*]Corresponding author

Figure 1: The frequency of tweeting behaviour of individuals in the months before, during and after an illness. Users significantly (check) decrease their rate of tweeting during the time that they had influenza. Dashed lines indicate the mean rate for the three months before / after the illness. (Todo: check significance)

4.3 Network Based Signals

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