

# Lecture Notes in Computer Science: Authors' Instructions for the Preparation of Camera-Ready Contributions to LNCS/LNAI/LNBI Proceedings

Alfred Hofmann

Springer-Verlag, Computer Science Editorial,  
Tiergartenstr. 17, 69121 Heidelberg, Germany  
{alfred.hofmann, ursula.barth, ingrid.haas, frank.holzwarth,  
anna.kramer, leonie.kunz, christine.reiss, nicole.sator,  
erika.siebert-cole, peter.strasser, lnsc}@springer.com  
<http://www.springer.com/lnsc>

**Abstract.** The abstract should summarize the contents of the paper and should contain at least 70 and at most 150 words. It should be written using the *abstract* environment.

**Keywords:** We would like to encourage you to list your keywords within the abstract section

## 1 Introduction

Social media has greatly democratized content creation. Facebook, Twitter, Skype, Whatsapp and LiveJournal are commonly used to share any thoughts and opinions about anything in the surrounding world. All this content has created new opportunities to study public opinion. Twitter is especially popular for research due to its scale, representative, variety of topics discussed and its ease access to content. In the past years, research in that direction was hindered by the unavailability of suitable datasets and lexicons for system training, development and testing. Some Twitter-specific resources were developed, but they were either small and proprietary, or they relied on noisy labels obtained automatically. This all changed with the shared task on Sentiment Analysis on Twitter, which is part of the international Workshop on Semantic Evaluation. The task is active since 2013 and it attract over 40+ participant teams. It contains 5 subtasks with their train and test datasets. First task is Message Polarity Classification, where we need to classify whether the given message is of positive, negative, or neutral sentiment. In second and third task, there is beside a given message, also a topic and you need to classify the message on two point scale and five-point scale, depending on the task. Forth and fifth task are also on two point and five point scale, where there are given tweets about a given topic, and you need to estimate their distribution.

## 2 Dataset

For tasks we are provided with their datasets with annotated tweets. They are gathered in a way, that express sentiment about popular topics. For this purpose, they extracted named entities from millions of tweets. The collected tweets were greatly skewed towards the neutral class. To reduce the class imbalance, they removed those that contained no sentiment-bearing words. Tweets are then manually filtered to obtain a set of meaningful topics with at least 100 tweets each. Topics that are ambiguous (e.q., Barcelona, which is a city or sport club) or too general (e.q., Paris). The topics in the training and in the test data do not overlap, meaning that test tweets consist of topics that are different from the topics in train dataset. Dataset is consisted of four parts: TRAIN (for training models), DEV (for tuning models), DEVTEST (for development-time evaluation) and TEST (for official evaluation). The first three datasets were annotated using Amazon's Mechanical Turk , while the TEST dataset was annotated on CrowdFlower.

### 2.1 Evaluation and discussion

#### 2.2 Subtask A: Message polarity classification

Subtask A is a single-label multi-class(SLMC) classification task. Each tweet must be classified as positive, neutral or negative. The evaluation score is measured as  $F_1^{PN}$ :

$$F_1^{PN} = \frac{F_1^P + F_1^N}{2} \quad (1)$$

$F_1^P$  is  $F_1$  for the positive class:

$$F_1^P = \frac{2\pi^P\varphi^P}{\pi^P + \varphi^P} \quad (2)$$

Here  $\pi^P$  and  $\varphi^P$  denote precision and recall for the positive class, respectively:

$$\pi^P = \frac{PP}{PP + PU + PN} \quad (3)$$

$$\varphi^P = \frac{PP}{PP + UP + NP} \quad (4)$$

where  $PP$ ,  $UP$ ,  $NP$ ,  $PU$ ,  $PN$  are the cells of the confusion matrix shown in table.

predicted/real	positive	neutral	negative
positive	PP	PU	PN
neutral	UP	UU	UN
negative	NP	NU	NN

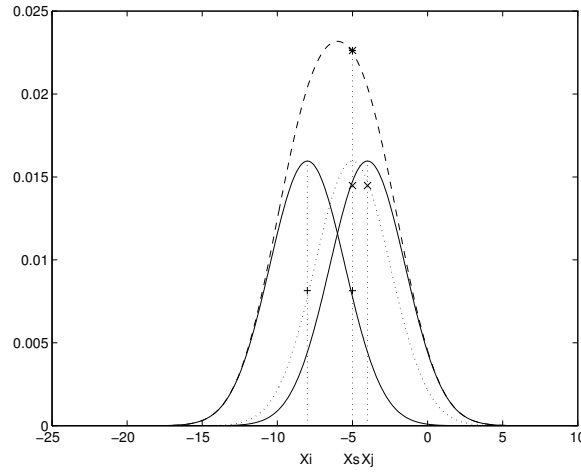
### 2.3 Subtask B: Tweet classification according to a two-point scale

For subtask B each tweet must be classified as either positive or negative. For this subtask they have adopted macro-averaged recall:

$$\varphi^{PN} = \frac{1}{2}(\varphi^P + \varphi^N) = \frac{1}{2}\left(\frac{PP}{PP + NP} + \frac{NN}{NN + PN}\right) \quad (5)$$

$\varphi^{PN}$  ranges from 0 to 1, where a value of 1 is the best score, while 0 means that classifier misclassified all items.

### 2.4 Figures



**Fig. 1.**

### 2.5 Formulas

$$\psi(u) = \int_o^T \left[ \frac{1}{2} (\Lambda_o^{-1}u, u) + N^*(-u) \right] dt. \quad (6)$$

## 3 The References Section

### References

1. Smith, T.F., Waterman, M.S.: Identification of Common Molecular Subsequences. J. Mol. Biol. 147, 195–197 (1981)

2. May, P., Ehrlich, H.C., Steinke, T.: ZIB Structure Prediction Pipeline: Composing a Complex Biological Workflow through Web Services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148–1158. Springer, Heidelberg (2006)
3. Foster, I., Kesselman, C.: The Grid: Blueprint for a New Computing Infrastructure. Morgan Kaufmann, San Francisco (1999)
4. Czajkowski, K., Fitzgerald, S., Foster, I., Kesselman, C.: Grid Information Services for Distributed Resource Sharing. In: 10th IEEE International Symposium on High Performance Distributed Computing, pp. 181–184. IEEE Press, New York (2001)
5. Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration. Technical report, Global Grid Forum (2002)
6. National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov>