Distinguishing Individuals from Organisations on Twitter

Sunghwan Mac Kim, Cécile Paris, Robert Power, Stephen Wan
Data61, CSIRO, Australia
{firstname.lastname}@csiro.au

ABSTRACT

Twitter is a popular public platform for individuals to share information and express opinions. It is also widely used as an active communication channel by many organisations. Analysis results about public opinion could thus be misrepresented and distorted if the posts generated by non-individuals are appropriately handled motivating the need to identify the type of user accounts for social media analytics. In this paper, we demonstrate the utility of a joint text and network representation for classifying if a profile belongs to an individual or organisation. Crucially, combining the two feature types has not been shown before for this task. Our experimental results show that performance gains are possible but only when using our novel adaptation of the Text-Associated DeepWalk (TADW) method.

Keywords

Twitter; User Classification; Social Network Embeddings

1. INTRODUCTION

Twitter is a popular public platform for individuals to share information and express opinions. As a result, Twitter has become a valuable large-volume and real-time resource for gauging public opinion on various social issues. In doing so, one has to be mindful of the various biases that the data might exhibit. For example, it has been shown that it is possible to exploit Twitter sampling techniques to maximise exposure [4], a method that can be utilised by spam bots and campaign managers alike. To ensure results obtained reflect public opinion, i.e., views of the individual members of the public, it is important to differentiate between posts generated by individuals versus posts authored by organisations. This is what we address in this work: Individual-vs-Organisation (IvO) classification, a crucial filter in addressing data bias.

Distinguishing between the two types of accounts is not trivial. For example, departments and companies are often represented on Twitter by a "public face", that is, a CEO or director. Such profiles may superficially appear similar to profiles for individuals but are really used to transmit the messages of the parent organisation and should really be treated as belonging to an organisation.

In this paper, we describe a text classification approach to tackle the problem of IvO classification. Intuitively, information from the profile text and the network structure could inform such a classifier. Prior work on IvO classification has examined these types of features, but only in isolation (e.g. text features [1], network [3]), without combining the two. Here, we show that it is desirable to do so using a joint representation of both text and network features, which, to our knowledge, has not been done before for IvO classification.

This paper describes a novel improvement upon one of the state-of-the-art methods for learning such a representation, Text-Associated DeepWalk (TADW) [6]. We outline our experimental results on held-out labelled data demonstrating that a joint representation obtained via our method outperforms others in distinguishing between profiles belonging to an individual versus an organisation.

2. A TEXT-NETWORK REPRESENTATION

To obtain a joint text and network low-dimensional representation (hereafter, text-network embedding or TNE) of a Twitter user profile, we adapt the TADW framework [6]. TADW is a leading framework capable of incorporating text content with DeepWalk, a network-oriented matrix factorisation method, producing TNE representations. However, the default TADW approach ultimately relies on an outmoded representation of text features, motivating the adaptation described below.

We define a Twitter profile network as $G = (P, R)$, where P is the set of profiles and R is the set of relationships between the profiles. Each relationship $r \in R$ is an unordered and unweighted pair $r = (u, v)$, where $u, v \in P$. Given G, we build a transition matrix T s.t. $T_{ij} = 1/d_i$ if $(i, j) \in R$ and $T_{ij} = 0$ otherwise, where d_i is the degree of a profile i. TADW requires T to be in the form $M \in \mathbb{R}^{|P|\times|P|}$, where $M = (T + T^2)/2$. M is factorised as: $M \rightarrow W^T \times H \times E$, where $W^T \in \mathbb{R}^{|P|\times k}$, $H \in \mathbb{R}^{k\times s}$, and $E \in \mathbb{R}^{s\times|P|}$, and where k and s are the number of dimensions for network and text embeddings, respectively. TADW produces W, the mapping from profiles to a reduced space k, and H, a mapping between the two reduced spaces, where E, the profile-to-text representation, is computed beforehand.

Originally, TADW computes E using Singular Value Decomposition (SVD) on a Term Frequency-Inverse Document Frequency (TF-IDF) weighted term-by-document matrix.

WWW 2017 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3054217

Table 1: Twitter profile identification results.

Approach	Precision	Recall	F1	Accuracy
Majority	0.208	0.333	0.256	0.625
MaxEnt with TNE	0.530	0.518	0.518	0.795
MaxEnt with Text	**0.641**	0.560	0.559	0.821
MaxEnt with Network	0.431	0.426	0.423	0.673
MaxEnt with Text + Network	0.551	0.549	0.546	0.784
Maxent with TNE-s2v	0.560	**0.576**	**0.568**	**0.851**

In our work, we modify TADW to use *sent2vec* [2] to learn E from the text contents of both labelled and unlabelled profiles, where each profile description is treated as a single text unit (a quasi-sentence). The sent2vec approach provides a low-dimensional (of size s) vector representation based on neural network approaches.

W and H are concatenated to form the joint representation TNE, which is used as a feature vector for each vertex (i.e., each profile). When sent2vec is used, we refer to this as the TNE-s2v representation. Notably, this representation can then be used for downstream applications. In this paper, it is used in a supervised machine learning scenario (here, we use a maximum entropy classifier), using a labelled dataset for IvO classification.

3. EXPERIMENTS

3.1 Dataset

We constructed a dataset of labelled Twitter profiles containing both network *and* text content, since, to our knowledge, no such datasets are available. Starting with a randomly sampled set of Twitter profiles, for each profile, we used the Twitter API to obtain followers up to a depth of 2 hops. In this dataset, $|P| = 3,017$ and $|R| = 38,785$. The profiles were annotated (using crowdsourcing) with the three labels (*individual*[1], *organisation*[2] and *other*[3]), with an inter-annotator agreement of 91.84%.

3.2 Results and Discussion

We use two baselines: a majority class baseline and the representation from the original TADW approach. For the machine learning results, we report precision, recall, F1 and accuracy scores based on 5-fold cross-validation and use a McNemar paired test with $p < 0.05$ for statistical significance. For all models, we set $k = 200$ and $s = 200$ (other values were trialled but did not improve performance). The evaluated models are described as follows:

- Majority: always predicts the most frequent label (namely, *individual*, 62.5%).
- MaxEnt with TNE: the original TADW approach using SVD for the profile-to-text representation.
- MaxEnt with Text: uses text embedding derived from the text contents of profiles.
- MaxEnt with Network: uses relationship embedding derived from the relationships between profiles.
- MaxEnt with Text + Network: uses the naively concatenated embedding of Text and Network relationships.
- Maxent with TNE-s2v: the joint embedding of Text and Network relationships.

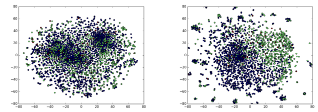
Table 1 shows the performance scores of our proposed models. Consistent with prior work, Text features lead to reasonable performance. We see that naively concatenating features for use in the maximum entropy classifier worsens performance. Even using the original formulation of the TADW method to obtain a joint TNE representation fails to improve F1 or accuracy. However, using our TNE-s2v representation leads to a statistically significant improvement in F1 and accuracy over the purely Text representation. This demonstrates that it is possible to capitalise on both text and network features for IvO classification but only with our adaptation.

To see the effectiveness of the joint representation over a naive concatenation of features in distinguishing between the individual and organisation classes, Figure 1 presents a 2-dimensional rendering of the representations using t-SNE [5] (selecting 200 dimensions for each embedding). We see that the TNE representation produces 2 visibly distinct clusters compared to the Text+Network representation.

[1]someone from the general public or a well-known person.

[2]A company, government department, non-profit organisation or its spokesperson.

[3]bots, people's pets, objects.



(a) T + N (b) TNE

Figure 1: Visualisation of Text+Network and TNE-s2v representations. Nodes represent Twitter profiles and colour denotes the profile types (blue:*individual*, green:*organisation*, brown:*other*).

4. CONCLUSIONS AND FUTURE WORK

In this paper, we tackled the problem of identifying individuals and organisations on Twitter, a step in addressing data sample bias for social media analytics. We demonstrated that capitalising on text and network features for this task leads to better results but only through a joint representation obtained via our novel adaptation of the TADW method. In future work, we intend to explore how this filtering mechanism can play a role in the so-called "post-truth" era to detect attempts by organisations to influence measures of public opinion.[4]

5. REFERENCES

[1] L. De Silva and E. Riloff. User type classification of tweets with implications for event recognition. *SDPASM*, 2014.

[2] Q. Le and T. Mikolov. Distributed representations of sentences and documents. *ICML*, 2014.

[3] J. McCorriston, D. Jurgens, and D. Ruths. Organizations are users too: Characterizing and detecting the presence of organizations on Twitter. *ICWSM*, 2015.

[4] F. Morstatter, H. Dani, J. Sampson, and H. Liu. Can one tamper with the sample api?: Toward neutralizing bias from spam and bot content. *WWW '16 Companion*, 2016.

[5] L. van der Maaten and G. E. Hinton. Visualizing high-dimensional data using t-sne. *JMLR*, 9, 2008.

[6] C. Yang, Z. Liu, D. Zhao, M. Sun, and E. Y. Chang. Network representation learning with rich text information. *IJCAI*, 2015.

[4]See for example, Google and Facebook's recent efforts to detect and highlight fake news: http://nyti.ms/2eUQIAT.