The Semantic Evolution of General and Specific Communities

Matthew Rowe[1] and Claudia Wagner[2]

[1] School of Computing and Communications, Lancaster University, Lancaster, UK
[2] Institute for Information and Communication Technologies, JOANNEUM RESEARCH, Graz, Austria
m.rowe@lancaster.ac.uk, claudia.wagner@joanneum.at

Abstract. Content injection methods rely on understanding community dynamics (i.e. attention factors) in order to publish content that community users will engage with (e.g. product-related posts), however such methods require re-training should the community's discussed topics change. In this paper we present an examination of the semantic evolution of community forums by measuring the topical specificity of online community forums and then tracking changes in the concepts discussed within the forums over time. Our results indicate that general discussion communities tend to diverge in their semantics, while topically-specific communities do not. These findings inform content injection methods on model longevity and the need for adaptation for general communities.

1 Introduction

The life cycles of online communities have been examined in [2], where the authors identified disparate evolution stages (e.g. *creation, growth, etc.*); and in [1], where Mozilla dev communities exhibited such stages. Although such works examine the changing nature of online communities, they do not examine how the topics of online communities evolve over time. Previously [3], we found a relation between the topical specificity of online communities and their attention patterns: single topic communities (e.g. Golf) require content to match this topic exactly. Therefore changes in a community's topics would require any models (e.g. content injection) that rely on the topical dynamics to be re-adapted. In this paper we examine the following question: *How is the topical specificity of an online community forum related to its evolution?* First, we describe a method to measure the topical specificity of community forums, before second, explaining how the community specificity is related to changes in discussed concepts.

2 Measuring Topical Specificity

For our analysis we used a dataset of 230 community forums from the Irish community message board Boards.ie.[1] Our method for measuring the topical

[1] http://www.boards.ie

P. Cimiano et al. (Eds.): ESWC 2013, LNCS 7955, pp. 277–278, 2013.

specificity of community forums took all posts published during a one-week window from 23/3/2005[2] and extracted entities using Zemanta.[3] We then used the DBPedia Ontology to identity the *type* of each entity to provide the set of concepts for a given forum: $A_f^{t't''}$. We measured the specificity of each community forum using different combinations of abstraction measures (Network Entropy, Degree Centrality, Eigenvector Centrality, Hits Authority/Hub Score, Statistical Subsumption, Key Player Problem) and composite functions (Most Specific Concept, Mean Specificity, Most Frequent Concept, Concept Frequency-Inverse Forum Frequency) that function over the DBPedia Ontology represented as a concept graph. For each model, we ranked the forums by their specificity scores and compared this to the ground truth ranking of forums ordered by hierarchical levels. We found the best model to be Concept Frequency with Eigenvector Centrality (Kendall $\tau_b = 0.075$), surpassing the random Knuth Shuffle baseline.

3 Semantic Evolution

To examine the relation between topical specificity and semantic evolution we divided the forums up into 10-equal frequency bins based on their specificity value (derived using the above model) and selected forums in the top and bottom bins to form *high* and *low*-specificity forums respectively. For these forums (23 in each bin) we randomly chose 4 weekly periods following 30/03/2005, extracted entities from forum posts during these periods, and derived concepts using the DBPedia Ontology. We then derived a concept frequency vector $\boldsymbol{c}$ for each forum for the four randomly-inspected analysis periods: $\{\boldsymbol{c}_1, \boldsymbol{c}_2, \boldsymbol{c}_3, \boldsymbol{c}_4\}$, and calculated the cosine similarity of the concept vectors between consecutive time periods (i.e. $cosine(\boldsymbol{c}_i, \boldsymbol{c}_{i+1})$). We examined the differences between the cosine similarity distributions of the *low* and *high* forums and found a significant difference in their means ($\mu_{low} = 0.783$ and $\mu_{high} = 0.854$ with $p < 0.05$ using Student's t-test). This result indicates that the initial specificity of a community forum is related to changes in concepts: *general communities exhibit greater semantic drift than specific communities*. Such results inform content injection models on the potential changes in topical dynamics and for model adaptation.

References

1. Hong, Q., Kim, S., Cheung, S.C., Bird, C.: Understanding a developer social network and its evolution. In: Proceedings of the 2011 27th IEEE International Conference on Software Maintenance, ICSM 2011, pp. 323–332. IEEE Computer Society, Washington, DC (2011)
2. Iriberri, A., Leroy, G.: A life-cycle perspective on online community success. ACM Comput. Surv. 41(2), 11:1–11:29 (2009)
3. Wagner, C., Rowe, M., Strohmaier, M., Alani, H.: Ignorance isn't bliss: an empirical analysis of attention patterns in online communities. In: AES Conference on Social Computing (2012)

[2] The median number of posts (4,455) were published on Boards.ie on this date.
[3] `http://www.zemanta.com`