Towards a Framework for Social Semiotic Mining

Eirini Giannakidou*, Athena Vakali*, Nikolaos Mavridis

*Aristotle University of Thessaloniki
◇ NCSR Democritos
Overview

Semiotics: From antiquity to Pierce and beyond!

Social Media & Data Mining

Our Goal: A Framework for:

a) Re-examining social content and tagging
b) Analyzing data mining on social content

Here, we will:

• Motivate & outline framework
• Apply to existing algorithms
• Demonstrate generalization to new
Outline

1. Introduction
2. Signs and Semiotics
3. Social Media as Semiotic Resources
4. Social Media Clustering
5. Discussion
6. Conclusion
1. INTRODUCTION
Setting

- Web 2.0
- Automated and Semi-Automated Content Analysis
- Wisdom of Crowds
- Applications: Recommender Systems, Policy Planning, Market Research etc.

- **However:** No clear theoretical framework!

*Image by: hermosaschool.wikispaces.com*
2. SIGNS AND SEMIOTICS
Signs

• “Nothing is a sign, unless it is interpreted as a sign” says Pierce
• Convention often non-conscious
• Natural Signs vs. Conventional Signs
• Symbol = “Συν” + “Βάλλω”
• Smoke + Fire
• But how could this apply to Social Media?
3. SOCIAL MEDIA AS SEMIOTIC RESOURCES
Ancient Beginnings

- Peri Hermineias (De Interpretazione) of Aristotle: Differentiated between objects, the words that refer to them and their corresponding experiences of the soul (psyche).

- also … Stoic Philosophers
The Semiotic Triangle

**THOUGHT OR REFERENCE**

- Symbolizes (a causal relation)
- Refers to (other causal relations)

**SYMBOL**

- Stands for (an imputed relation)

**REFERENT**

Ogden & Richards
Extended Semiotic Triangle

Carter, B., & Knight, D.
Versions exist by Pierce, Ullman, Harnad, Vogt etc.
Transfer to Social Web

THOUGHT OR REFERENCE

SYMBOL → REFERENT

TAG → WEB 2.0 RESOURCE

Transfer to Social Web
The three spaces

- Tag space
- Concept space
- Resource space
Ways to search for relations between entities (I)

1) Examination of first level relationships (taking into account entities in one vertex only)

(a') Entities of the same vertex
Ways to search for relations between entities (II)

2) Examination of second level relationships (taking into account entities in two vertices)

(β') Entities of multiple vertices
3) Extension of the semiotic triangle across multiple axis (for example across the temporal axis or the user axis)
4. SOCIAL MEDIA CLUSTERING
Generalized Social Clustering Framework (I)

**Step 1)** Chose type of clustering: One-way clustering (L1), co-clustering (L2)

**Step 2)** Determine which subset of vertices $U$ from $V=(T,C,R)$ will be used for the distance function for the clustering, i.e.

- 1 chosen vertex set out of $V$ for the case of one-way clustering (L1), i.e. $U=(V_1)$ where $V_1$ belongs to $V$ (T, C, or R)
- 2 chosen vertex sets out of $V$ for the case of co-clustering (L2), i.e. $U=(V_1,V_2)$ where $V_1$ and $V_2$ belong to $V$ but $V_1$ not equal to $V_2$
Generalized Social Clustering Framework (II)

**Step 3)** Form similarity spaces within each vertex of the semiotic triangle: introduce symmetric similarity/distance metrics, one for each vertex of the semiotic triangle.

**Step 4)** Introduce transformation mappings across the three vertices of the semiotic triangle, i.e. \( T(R), T(C), R(T), R(C), C(T), C(R) \), where for example \( C(T) \) refers the set of concepts \( C \) that is related to a specific tag or set of tags \( T \).
Generalized Social Clustering Framework (III)

**Step 5)** Introduce generalized distances, depending on whether the case is L1 or L2:

Example:
\[ d_G(V_1,V_1) = w_1 d V_1(V_1,V_1) + w_2 d V_2(V_2,V_2) + w_3 d V_3(V_3,V_3) \] (1)

where \( V_2 = V_2(V_1) \), i.e. the set of entities belonging to vertex \( V_2 \) that arise out of the transformation mapping \( V_2(V_1) \).

e.g. if \( V_2 = C \) (concepts) and \( V_1 = T \) (tags),

then \( V_2(V_1) = C(T) = \) the concepts that correspond to tag \( T \)
then, proceed by clustering according to the distance \( d_G(V_1,V_1) \)
Application to previous work


• Bumgardner J. 2006. Experimental colr pickr. Available at: http://www.krazydad.com/


• Anastasia Stampouli, Eirini Giannakidou, and Athena Vakali. 2010. Tag disambiguation through Flickr and Wikipedia. In Proceedings of the 15th international conference on Database systems for advanced applications (DASFAA’10),
Application to previous work

• another case L2 (co-clustering) where 2 vertices are used for the co-clustering: tags and resources (artists) [26].... numerous other such examples exist.

.... **moving beyond** single-vertex and dual-vertex one way clustering, and also beyond co-clustering, there exist methods that extend the **semiosis** across users and across the temporal axis. For example, although in [18] the first stage of clustering uses tags only, at the second stage the user axis is utilized. In numerous other papers the temporal axis is also taken into account for clustering [28].
5. DISCUSSION
Discussion

• Existing cases covered by framework;
• One extend to novel cases:
  can also produce novel combinations
that fall within the generative power of the framework.

For example, one could create novel methods by choosing appropriate similarity metrics within each vertex, choosing subsets of vertices in order to create generalized weighted distances that contain similarities arising across more than one vertex (for example, one could use the triple combination tags – concepts – auditory features of resources), and perform either one-way clustering, or co-clustering, or even extend to higher-dimensional tensor-based methods.

Future extensions to framework:
- better treatment of user and temporal axis
- Move beyond clustering to classification and regression
Towards a Framework for Social Semiotic Mining

Eirini Giannakidou, Athena Vakali, Nikolaos Mavridis

THANK YOU!