Mining Opinion Attributes From Texts using Multiple Kernel Learning

Aleksander Wawer
axw@ipipan.waw.pl

December 11, 2011

Institute of Computer Science
Polish Academy of Science

Agenda

- Text Corpus
- Ontology Relations
- Complex Attributes
- Similarity Matrices and Kernels
- MKL
- Results

Text Corpus

- The reviews are a random sample from `http://cokupic.pl`, the largest Polish opinion aggregator service.
- Texts annotated manually by 4 annotators, who first indentified attributes by reading texts, then annotated the texts using these attributes. Finally, attributes and annotations were unified into a single set of annotations and an ontology.
- Two product types: perfume (4232 words) and women underwear (2446 words). Each product type had a corresponding set of attributes. The number of attributes: 20 for underwear and 31 for perfume.
- The numbers of annotated text fragments containing references to attributes: 305 and 343. Perfume reviews are more wordy.

Ontology Relations

The structure of attributes was found to be hierarchical, with the root being the evaluated object (product type). Annotators identified following relationships between attributes:

- **is associated with** . The most general relation which does not specify exact semantics which links both attributes.
- **is part of** : meronymy. Eg., the attribute *bottle* may be linked using this type of relation with the attribute *container*
- **is property of** : an attribute is a feature that describes a higher level attribute, for example perfume *durability* may be a property of *fragrance*.
- **is type of** : an attribute as a type of another, higher level attribute. For example, attribute perfume attribute *special occasion* is a type of *purpose*.

We flatten the tree and treat attributes as independent.

Attribute Complexity

Typically: attributes, marked in [..], are expressed by nouns and evaluative concepts, marked in (..), by adjectives:

```
[odswiezajacy] i [energizujacy] <zapach>
[refreshing] and [energizing] <fragrance>
```

This leads to attribute-value pairs extraction as in Hu (2004). Especially in the perfume reviews reviewers identified many fragments like:

```
Te perfumy [<przyciagaja kobieca uwage>]
This fragrance [<attracts female attention>]
```

Attribute: “attracting attention of appropriate group of people”. Issues: (1) various syntactic formations; (2) verbs other than “attract”; (3) refer to different groups of people.

Similarity Measures - WordNet

- Resnick (1995) similarity is defined as follows:

$$res(c_1, c_2) = IC(LCS(c_1, c_2)))$$

$$IC(c) = -logP(c)$$

  where $P(c)$ – probability of a synset c in word frequency list

- Lin (1998) similarity:

$$lin(c_1, c_2) = 2\frac{res(c_1, c_2)}{IC(c_1) + IC(c_2)}$$

- Jiang and Conrath similarity (1997):

$$jcn(c_1, c_2) = \frac{1}{IC(c_1) + IC(c_2) + 2res(c_1, c_2)}$$

- Path similarity $path(c_1, c_2)$ is the shortest path distance between c_1 and c_2 in hypernym taxonomy.

Similarity Measures - Lexical

- *orth* – the number of identical words in their exact (ortographical) forms, common between two compared text fragments.
- *base* – the number of words whose base forms, obtained with morphological analysis and disambiguation, are identical between two compared text frames.

Kernel Matrices Visualization

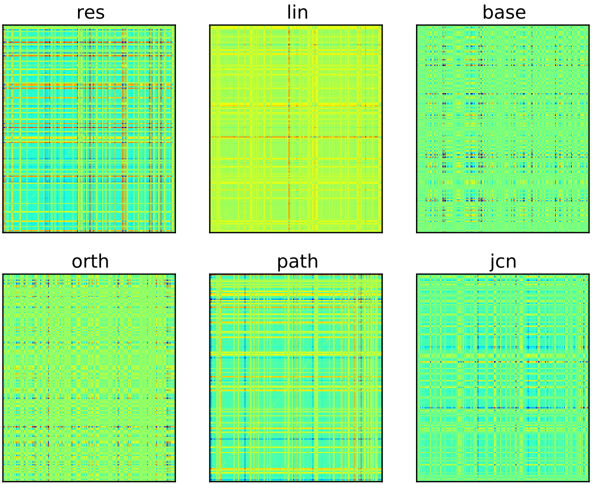

Figure: Kernels for the two product types.

From Similarities to Kernels

Two basic approaches (Chen et al., 2009):

- similarities as distances in some Euclidean space, often based on multidimensional scaling (MDS)
- similarities as inner products in some Hilbert space, often using support vector machine (SVM)

The formulation of SVM is as follows:

$$\begin{aligned} \textit{maximize}: \quad & 1^\top \alpha - \frac{1}{2}\alpha^\top \textit{diag}(y) K \textit{diag}(y) \alpha \\ \textit{subject to}: \quad & 0 \preceq \alpha \preceq C, y^\top \alpha = 0 \end{aligned} \tag{1}$$

where $C > 0$ is a hyperparameter, K a kernel matrix whose (i, j)-entry is $K(x_i, x_j)$.
Theory: SVM's convex optimization demands the corresponding kernel matrix K to be positive semidefinite (PSD).

Multiple Kernel Learning (MKL)

Following Sonnenburg et al. (2006): a semi-infinite linear program, solved by a standard LP solver and an SVM implementation. The algorithm seeks optimal linear combination of M kernels:

$$k_{combined}(x, x') = \sum_{m=1}^{M} \beta_m k_m(x, x')$$

- Each cross validation run resulted in slightly different weight (β) distributions;
- Generally: *res* and *base* dominated over others, β_{res} at 0.2 - 0.3, β_{base} 0.6 - 0.7. Optimal results were obtained by mixtures of semantic and lexical similarities.

Results

Avg accuracy in 5-fold cross validation. Training and testing partition sizes set to 90% and 10%.

Random baseline accuracy is at 0.05 for underwear reviews (20 attributes) and 0.03 for perfume reviews (31 attributes).

	Underwear		Perfume	
	acc.	C	acc.	C
res	0.31	0.001	0.42	0.1
lin	0.36	0.001	0.44	0.001
base	0.43	0.5	0.51	0.001
orth	0.36	1.0	0.47	0.1
path	0.39	0.5	0.44	0.1
jcn	0.3	0.001	0.43	0.1
MKL	0.46	5.0	0.53	5.0

Table: Average accuracies (acc.) and corresponding C parameter values for each similarity matrix and MKL matrices.

Summary

1. A novel framework for recognizing complex opinion attributes from product reviews.
2. Instead of focusing on linguistic properties of text fragments and their direct representations, we focus on fragments' similarities using multiple sources of lexical and semantic information.
3. Problem: multi-class classification, multiple similarity matrices.
4. Solution: multiple kernel learning algorithm seeks optimal combinations of matrices using linear programming and support vector machines for classification.
5. Experiments demonstrate benefits from multiple sources of information.
6. The approach is promising especially for reviews of product types with complex and wordy attribute expressions.