A Knowledge-based Semantic Kernel for Text Classification

Jamal Abdul Nasir
e-mail: jamaln@lums.edu.pk
LUMS, Pakistan

Asim Karim
e-mail: akarim@lums.edu.pk
LUMS, Pakistan

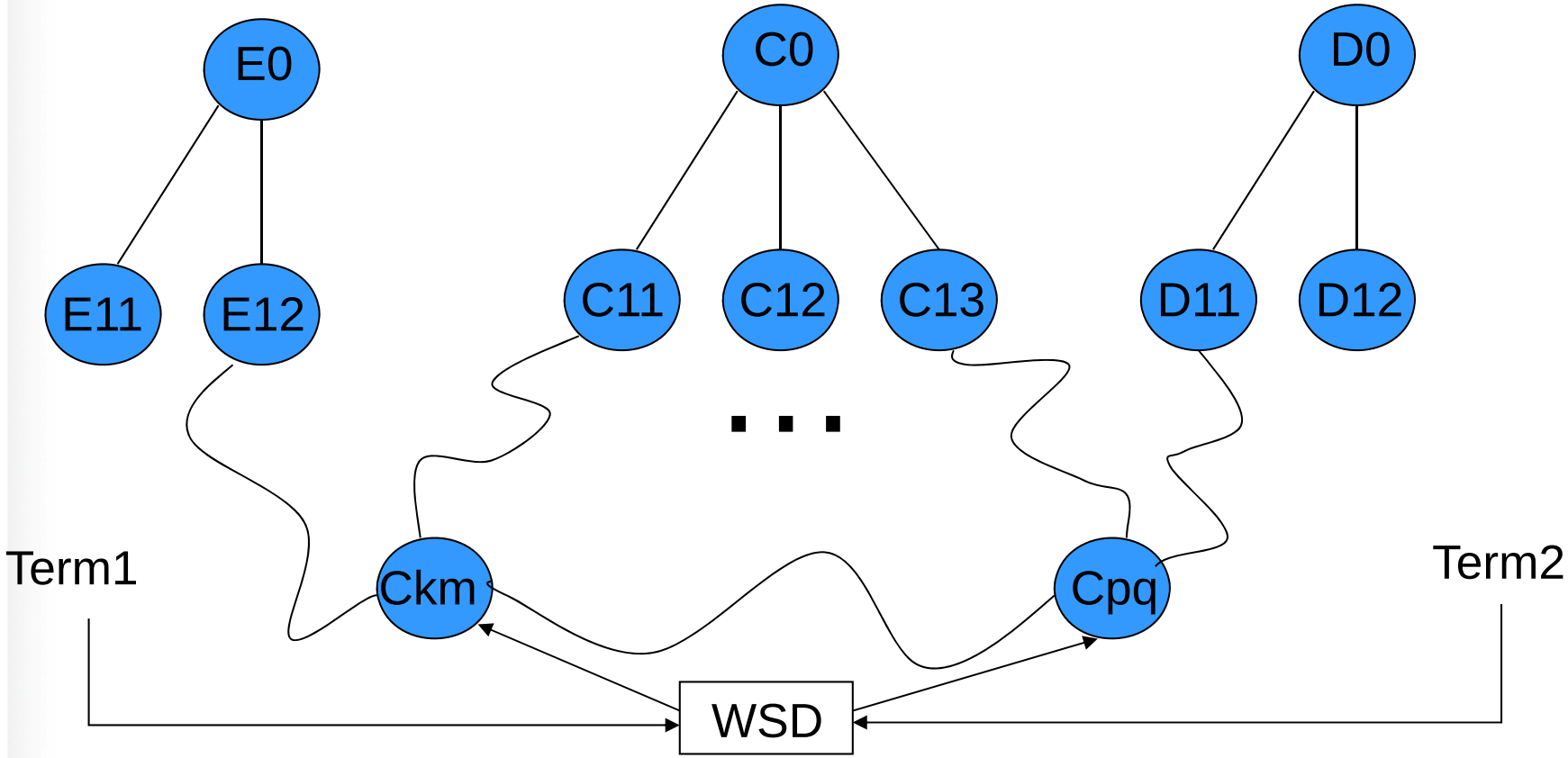
George Tsatsaronis
e-mail: george.tsatsaronis@biotec.tu-dresden.de
BIOTEC, Germany

Iraklis Varlamis
e-mail: varlamis@hua.gr
HUA, Greece

Introduction - Motivation

- Typical textual representation is *BOW (“Bag of Words Representation”)*
  - **Synonymy** of terms – affects **recall**
  - **Polysemy** of terms – affects **precision**
  - Semantic similarity between different terms is not taken into account
- Semantic Kernel for Text Classification
  - Based on a measure of Semantic Relatedness (OMIOTIS)
  - Combines semantic similarity and surface string matching
  - Uses a knowledge base – the *WordNet* thesaurus for the English language
  - Can be embedded in any classifier

Background Information – *Omiotis* Measure of Semantic Relatedness



Background Information – *Omiotis* Measure of Semantic Relatedness

$$SCM(S,O) = \prod_{i=1}^{l} e_i$$

$$SPE(S,O) = \prod_{i=1}^{l} \frac{2d_i d_{i+1}}{d_i + d_{i+1}}$$

$$\max(SCM(S,O) \bullet SPE(S,O))$$

Semantic Kernel using *Omiotis*

$$\boldsymbol{\varphi}(d) = (\boldsymbol{\varphi}(d)^T R)^T$$

$$\kappa(d_i, d_j) = \boldsymbol{\varphi}(d_i)^T \boldsymbol{\varphi}(d_j) = \boldsymbol{\varphi}(d_i)^T RR^T \boldsymbol{\varphi}(d_j)$$

Experimental Setup

- Four text classification data sets
  - **MoviewReview:** 2 classes, 2,000 documents (1,000 each class)
  - **Ohsumed.91:** 15 classes, Medline documents for 1991
  - **20 Newsgroups:** 20 categories, approximately 20,000 documents
  - **WebKB:** 7 classes, approximately 8,000 documents
- Four classifiers used
  - **Support Vector Machines** (SVM)
  - **Naïve Bayes** (NB)
  - **Maximum Entropy** (ME)
  - **Balanced Winnow** (BW)
- 10-fold cross validation, average accuracy measured

Experimental Evaluation

	MovieReview	Ohsumed	20 Newsgroups	WebKB
SVM	83.30	55.15	90.08	86.37
$SVM_{Omiotis}$	**91.97**	**57.17**	**92.93**	84.58
NB	77.41	50.32	87.27	84.17
$NB_{Omiotis}$	84.13	51.29	90.44	88.52
ME	79.11	51.47	85.31	91.02
$ME_{Omiotis}$	81.86	50.17	87.35	**91.52**
BW	76.23	50.93	81.66	81.42
$BW_{Omiotis}$	79.25	51.83	84.58	85.34

Text classification performance – Accuracy in %

Advantages and Limitations

Semantic Kernel Advantages

- Terms that in BOW were not related, now, if they are semantically similar, the similarity is taken into account
- Semantic relations are coming from a reviewed source (WordNet)
- Computationally fast, if term-to-term relatedness values are precomputed
- Applicable to any classifier, replacing the similarity measure between instances (points)

Semantic Kernel Limitations

- Coverage of terms is bounded by the coverage of the used knowledge-base (in our case WordNet)
- Document-to-Document similarity needs more time, as the overlap between semantic related terms increases, compared to the overlap in the case of BOW (overlap only when exact match of terms occurs)

Conclusions – Future Work

- Semantic kernel improves BOW performance in text classification
  - Evaluation with four classifiers, in four data sets
  - Improvement up to 8.5 p.p.

- Future Work
  - Embed more knowledge sources, e.g., YAGO, to improve coverage
  - Evaluate more measures of semantic relatedness/similarity using the same kernel trick
  - Apply to more text mining tasks, e.g., clustering , document annotation (can be seen as classification), paraphrase detection

Thank you very much for your attention!

Questions / Comments ?