

Sentence-Level Subjectivity Detection Using Neuro-Fuzzy Models

Samir Rustamov

Georgia Institute of Technology
225 North Avenue NW
Atlanta, GA 30332, USA
samir.rustamov@gmail.com

Mark A. Clements

Georgia Institute of Technology
225 North Avenue NW
Atlanta, GA 30332, USA
clements@ece.gatech.edu

Abstract

In this work, we attempt to detect sentence-level subjectivity by means of two supervised machine learning approaches: a Fuzzy Control System and Adaptive Neuro-Fuzzy Inference System. Even though these methods are popular in pattern recognition, they have not been thoroughly investigated for subjectivity analysis. We present a novel “Pruned ICF Weighting Coefficient,” which improves the accuracy for subjectivity detection. Our feature extraction algorithm calculates a feature vector based on the statistical occurrences of words in a corpus without any lexical knowledge. For this reason, these machine learning models can be applied to any language; i.e., there is no lexical, grammatical, syntactical analysis used in the classification process.

1 Introduction

There has been a growing interest, in recent years, in identifying and extracting subjective information from Web documents that contain opinions. Opinions are usually subjective expressions that describe people's sentiments, appraisals, or feelings. Subjectivity detection seeks to identify whether the given text expresses opinions (subjective) or reports facts (objective) (Lin et al., 2011). Automatic subjectivity analysis methods have been used in a wide variety of text processing and natural language applications. In many natural language processing tasks, subjectivity detection has been used as a first phase of filtering to generate more informative data.

The goal of our research is to develop learning methods to create classifiers that can distinguish subjective from objective sentences. In this paper,

we achieve sentence-level subjectivity classification using language independent feature weighting. As a test problem, we employed a subjectivity database from the "Rotten Tomatoes" movie reviews (see <http://www.cs.cornell.edu/people/pabo/movie-review-data>).

We present two supervised machine learning approaches in our development of sentence-level subjectivity detection: Fuzzy Control System (FCS), and Adaptive Neuro-Fuzzy Inference System (ANFIS). Even though these methods are popular in pattern recognition, they have not been thoroughly investigated for subjectivity analysis. We present a novel “Pruned ICF Weighting Coefficient,” which improves the accuracy for subjectivity detection. Our feature extraction algorithm calculates a feature vector based on statistical occurrences of words in the corpus without any lexical knowledge. For this reason, the machine learning models can be applied to any language; i.e., there is no lexical, grammatical, syntactical analysis used in the classification process.

2 Related work

In recent years, several different supervised and unsupervised learning algorithms were investigated for defining subjective information in text or speech.

Riloff and Wiebe (2003) presented a bootstrapping method to learn subjectivity classifiers from a collection of non-annotated texts. Wiebe and Riloff (2005) used a similar method, but they also learned objective expressions apart from subjective expressions.

Pang and Lee (2004) proposed a MinCut based algorithm to classify each sentence as being subjective or objective. The goal of this research was to remove objective sentences from each review to improve document-level sentiment classification (82.8% improved to 86.4%).

Grefenstette et al. (2004) presented a Web mining method for identifying subjective adjectives.

Wilson et al. (2004) and Kim et al. (2005) presented methods of classifying the strength of opinion being expressed in individual clauses (or sentences).

Riloff et al. (2006) defined subsumption relationships among unigrams, n -grams, and lexico-syntactic patterns. They found that if a feature is subsumed by another, the subsumed feature is not needed. The subsumption hierarchy reduces a feature set and reduced feature sets can improve classification performance.

Raaijmakers et al (2008) investigated the use of prosodic features, word n -grams, character n -grams, and phoneme n -grams for subjectivity recognition and polarity classification of dialog acts in multiparty conversation. They found that for subjectivity recognition, a combination of prosodic, word-level, character-level, and phoneme-level information yields the best performance and for polarity classification, the best performance is achieved with a combination of words, characters and phonemes.

Murray and Carenini (2009) proposed to learn subjective patterns from both labeled and unlabeled data using n -gram word sequences with varying level of lexical instantiation. They showed that learning subjective trigrams with varying instantiation levels from both annotated and raw data can improve subjectivity detection and polarity labeling for meeting speech and email threads.

Martineau and Finin (2009) presented Delta TFIDF, an intuitive general purpose technique, to efficiently weight word scores before classification. They compared SVM Difference of TFIDFs and SVM Term Count Baseline results for subjectivity classification. As a result, they showed that SVM based on Delta TFIDF gives high accuracy and low variance.

Barbosa and Feng (2010) classified the subjectivity of tweets (postings on Twitter) based on two kind of features: meta-information about the words on tweets and characteristics of how tweets are written.

Yulan He (2010) proposed subjLDA for sentence-level subjectivity detection by modifying the latent Dirichlet allocation (LDA) model through adding an additional layer to model sentence-level subjectivity labels.

Benamara et al. (2011) proposed subjectivity classification at the segment level for discourse-based sentiment analysis. They classified each segment into four classes, S, OO, O and SN, where S segments are segments that contain explicitly lexicalized subjective and evaluative expressions, OO segments are positive or negative opinion implied in an objective segment, O segments contain neither a lexicalized subjective term nor an implied opinion, SN segments are subjective, though non-evaluative, segments that are used to introduce opinions.

Remus (2011) showed that by using readability formulae and their combinations as features in addition to already well-known subjectivity clues leads to significant accuracy improvements in sentence-level subjectivity classification.

Lin et al, (2011) presented a hierarchical Bayesian model based on latent Dirichlet allocation, called subjLDA, for sentence-level subjectivity detection, which automatically identifies whether a given sentence expresses opinion or states facts.

All the aforementioned work focused on English data and most of them used an English subjectivity dictionary. Recently, there has been some work on subjectivity classification of sentences in Japanese (Kanayama et al., 2006), Chinese (Zagibalov et al., 2008; Zhang et al., 2009), Romanian (Banea et al., 2008; Mihalcea et al., 2007), Urdu (Mukund and Srihari, 2010), Arabic (Abdul-Mageed et al., 2011) and others based on different machine learning algorithms using general and language specific features.

Mihalcea et al., (2007) and Banea et al., (2008) investigated methods to automatically generate resources for subjectivity analysis for a new target language by leveraging the resources and tools available for English. Another approach (Banea et al., 2010) used a multilingual space with meta classifiers to build high precision classifiers for subjectivity classification.

Recently, there has been some work focused on finding features that can be applied to any language. For example, Mogadala and Varma (2012) presented sentence-level subjectivity classification using language independent feature weighting and performed experiments on 5 different languages including English and a South Asian language (Hindi).

Rustamov et. al., (2013) applied hybrid Neuro-Fuzzy and HMMs to document level sentiment analysis of movie reviews.

In the current work, our main goal is to apply supervised methods based on language independent features for classification of subjective and objective sentences.

3 Feature Extraction

Most language independent feature extraction algorithms are based on the presence or occurrence statistics within the corpus. We describe such an algorithm which is intuitive, computationally efficient, and does not require either additional human annotation or lexical knowledge.

We use a subjectivity dataset 1v.0: 5000 subjective and 5000 objective processed sentences in movie reviews [Pang/Lee ACL 2004].

As our target does not use lexical knowledge, we consider every word as one code word. In our algorithm we do not combine verbs in different tenses, such as present and past ("decide" vs "decided") nor nouns as singular or plural ("fact" vs "facts"). Instead, we consider them as the different code words.

Below, we describe some of the parameters:

- N is the number of classes (in our problem $N=2$: i.e. subjective and objective classes);
- M is the number of different words (terms) in the corpus;
- R is the number of observed sequences in the training process;
- $O^r = \{o_1^r, o_2^r, \dots, o_{T_r}^r\}$ are the sentences in the training dataset, where T_r is the length of r -th sentence, $r = 1, 2, \dots, R$;
- $\mu_{i,j}$ describes the association between i -th term (word) and the j -th class ($i = 1, \dots, M; j = 1, 2, \dots, N$);
- $c_{i,j}$ is the number of times i -th term occurred in the j -th class;
- $t_i = \sum_j c_{i,j}$ denotes the occurrence times of the i -th term in the corpus;
- frequency of the i -th term in the j -th class

$$\bar{c}_{i,j} = \frac{c_{i,j}}{t_i};$$

We present a new weighting coefficient, which affects the accuracy of the system, so that instead

of the number of documents we take the number of classes in the well-known IDF (Inverse-Document Frequency) formula. Similar to IDF, we call it Pruned ICF (Inverse-Class Frequency)

$$ICF_i = \log_2 \left(\frac{N}{dN_i} \right),$$

where i is a term, dN_i is the number of classes containing the term i , which $\bar{c}_{i,j} > q$, where

$$q = \frac{1}{\delta \cdot N}.$$

The value of δ is found empirically with $\delta = 1.4$ being best for the corpus investigated.

The membership degree of the terms ($\mu_{i,j}$) for appropriate classes can be estimated by experts or can be calculated by analytical formulas. Since a main goal is to avoid using human annotation or lexical knowledge, we calculated the membership degree of each term by an analytical formula as follows ($i = 1, \dots, M; j = 1, 2, \dots, N$):

$$\text{TF: } \mu_{i,j} = \frac{\bar{c}_{i,j}}{\sum_{v=1}^N \bar{c}_{i,v}}; \quad (1)$$

$$\text{TF} \cdot \text{ICF: } \mu_{i,j} = \frac{\bar{c}_{i,j} \cdot ICF_j}{\sum_{v=1}^N \bar{c}_{i,v} \cdot ICF_v}; \quad (2)$$

4 Subjectivity detection using Fuzzy Control System

We use a statistical approach for estimation of the membership function, instead of expert knowledge, at the first stage. Then we apply fuzzy operations and modify parameters by the back-propagation algorithm.

We now introduce our algorithm ($r = 1, 2, \dots, R$).

1. The membership degree of terms ($\mu_{i,j}^r$) of the r -th sentence are calculated from formulas (1)-(2).
2. Maximum membership degree is found with respect to the classes for every term of the r -th sentence

$$\begin{aligned} \bar{\mu}_{i,j}^r &= \mu_{i,j}^r, \\ j &= \arg \max_{1 \leq v \leq N} \mu_{i,v}^r, \\ i &= 1, \dots, M. \end{aligned} \quad (3)$$

3. Means of maxima are calculated for all classes:

$$\begin{aligned}\bar{\mu}_j^r &= \frac{\sum_{k \in Z_j^r} \bar{\mu}_{k,j}^r}{T_r}, \\ Z_j^r &= \left\{ i : \bar{\mu}_{i,j}^r = \max_{1 \leq v \leq N} \mu_{i,v}^r \right\} \\ j &= 1, \dots, N.\end{aligned}\quad (4)$$

We use the Center of Gravity Defuzzification (CoGD) method for the defuzzification operation.

Objective and subjective sentences selected according to classes are trained by a fuzzy control model. The objective function is defined as follows (Aida-zade et. al, 2012):

$$E(y) = \frac{1}{2} \sum_{r=1}^R \left(\frac{\sum_{j=1}^N \bar{\mu}_j^r y_j}{\sum_{j=1}^N \bar{\mu}_j^r} - d_r \right)^2 \rightarrow \min_{y \in R^N}, \quad (5)$$

$y = (y_1, y_2, \dots, y_N)$, $d_r \in \{1, 2, \dots, N\}$ desired output.

The partial derivatives of this function are calculated in following form:

$$\frac{\partial E(y)}{\partial y_t} = \sum_{r=1}^R \frac{\bar{\mu}_t^r}{\sum_{j=1}^N \bar{\mu}_j^r} \left(\frac{\sum_{j=1}^N \bar{\mu}_j^r y_j}{\sum_{j=1}^N \bar{\mu}_j^r} - d_r \right), \quad t = 1, 2, \dots, N.$$

Function (5) is minimized by the conjugate gradient method with the defined optimal values of y^* .

Rounding of \bar{y} shows the index of the classes obtained in the result:

$$\bar{y} = \frac{\sum_{j=1}^N \bar{\mu}_j y_j^*}{\sum_{j=1}^N \bar{\mu}_j}. \quad (6)$$

Acceptance strategy (s):

$$s = \begin{cases} i_s \in I, & \text{if } \bar{y} \in (i_s - \Delta_1, i_s + \Delta_1) \\ \text{reject,} & \text{otherwise} \end{cases},$$

where i_s is the index of the appropriate class, $I = \{1, 2, \dots, N\}$. Here $\Delta_1 \in [0; 0.5]$ is the main quantity, which influences the reliability of the system.

It is straightforward to check which feature vector gives the best results for FCS. Table 1. shows average accuracy over 10 fold cross validation of FCS based on (1)-(2) features in the non-restricted case. Note that these results depend on the classification method these results might be different for different classifiers.

Features	Accuracy (%)
TF	89.87
TF · ICF	91.3

Table 1. Results of FCS based on TF and TF · ICF features.

We also checked FCS based on Delta TFIDF features (Martineau and Finin, 2009). As DeltaIDF weighting coefficients of both classes are the same, application of DeltaIDF weighting does not change the accuracy of the FCS. As we see from Table 1., the accuracy of the method increases after application of Pruned ICF weighting.

We show results of subjectivity detection by FCS with different values of Δ_1 based on TF · ICF in Table 2. It can be seen that the rejection percentage is 0.01 for $\Delta_1 = 0.5$. In the testing process 0.01% of the sentences have such words, which after pruned ICF weighting, becomes 0 and the system rejects such sentences.

	Correct (%)	Rejection (%)	Error (%)
$\Delta_1 = 0.3$	76.41	20.86	2.73
$\Delta_1 = 0.4$	85.11	10.14	4.75
$\Delta_1 = 0.5$	91.3	0.01	8.69

Table 2. Average results of 10 folds cross validation accuracy of FCS based on TF · ICF feature with different value of Δ_1 .

5 Subjectivity detection using Adaptive Neuro Fuzzy Inference System

Fig. 1 illustrates the general structure of Adaptive Neuro Fuzzy Inference System. In response to linguistic statements, the fuzzy interface block provides an input vector to a Multilayer Artificial Neural Network (MANN) (Fuller, 1995).

We used statistical estimation of membership degree of terms by (2) instead of linguistic statements at the first stage. Then we applied fuzzy operations (3) and (4).

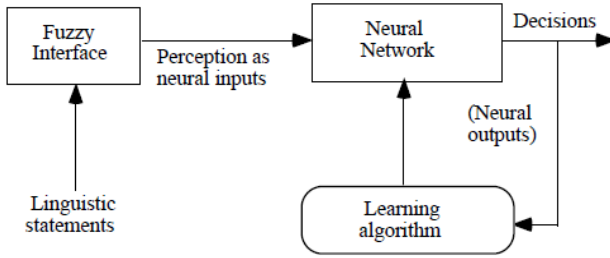


Fig. 1. The structure of ANFIS.

MANN was applied to the output of the fuzzyfication operation. The input vector of neural network is taken from the output vector of the fuzzyfication operation (fig. 2). Outputs of MANN are taken as indexes of classes appropriate to the sentences. MANN is trained by the back-propagation algorithm.

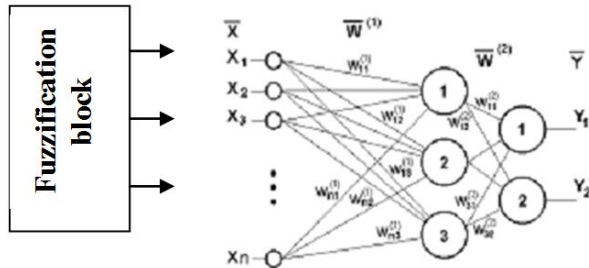


Fig. 2. The structure of MANN in ANFIS.

We set two boundary conditions for the acceptance decision:

- 1) $\bar{y}_k \geq \Delta_2$,
- 2) $\bar{y}_k - \tilde{y}_p \geq \Delta_3$,

where y is the output vector of MANN, \bar{y}_k and \tilde{y}_p are two successive maximum elements of the vector y , i.e.

$$\bar{y}_k = \max_{1 \leq i \leq N} y_i, \quad k = \arg \max_{1 \leq i \leq N} y_i,$$

$$\tilde{y}_p = \max_{1 \leq i \leq k-1; k+1 \leq i \leq N} y_i.$$

There is shown results of subjectivity detection in movie reviews by ANFIS with different values of Δ_2 and Δ_3 in Table 3.

	Correct (%)	Rejection (%)	Error (%)
$\Delta_2 = 0.8; \Delta_3 = 0.5$	78.66	18.84	2.5
$\Delta_2 = 0.5; \Delta_3 = 0.5$	85.77	8.62	5.61
No restriction	91.66	0.01	8.33

Table 3. Average results of 10 folds cross validation accuracy ANFIS based on TF·ICF for subjectivity detection in movie reviews.

The accuracy of the ANFIS (91.66%) is higher than that of FCS (91.3%) at the cost of additional variables being required in the middle layer of the neural network.

6 Conclusion

We have described two different classification system structures, FCS, ANFIS, and applied them to sentence-level subjectivity detection in a movie review data base. We have specifically shown how to train and test these methods for classification of sentences as being either objective or subjective. A goal of the research was to formulate methods that did not depend on linguistic knowledge and therefore would be applicable to any language. An important component of these methods is the feature extraction process. We focused on analysis of informative features that improve the accuracy of the systems with no language-specific constraints. As a result, a novel "Pruned ICF Weighting Function" was devised with a parameter specifically estimated for the subjectivity data set.

When comparing the current system with others, it is necessary to emphasize that the use of linguistic knowledge does improve accuracy. Since we do not use such knowledge, our results should only be compared with other methods having similar constraints, such as those which use features based on bags of words that are tested on the same data set. Examples include studies by Pang and Lee (2004) and Martineau and Finin (2009). Pang and Lee report 92% accuracy on sentence-level subjectivity classification using Naive Bayes classifiers and 90% accuracy using SVMs on the same data set. Martineau and Finin (2009) reported 91.26% accuracy using SVM Difference of TFIDFs. The currently reported results: FCS (91.3%), ANFIS (91.7%) are similar. However, our presented methods have some advantages. Because the function (5) is minimized only with respect to $y = (y_1, y_2, \dots, y_N)$ (in the defined problem $N=2$), FCS is the fastest algorithm among supervised machine learning methods. At the cost of additional variables added within the middle layer of the neural network, ANFIS is able to improve accuracy a

small amount. It is anticipated that when IF-THEN rules and expert knowledge are inserted into ANFIS and FCS, accuracy will improve to a level commensurate with human judgment.

References

- Aditya Mogadala, Vasudeva Varma. 2012. Language Independent Sentence-Level Subjectivity Analysis with Feature Selection. *Proceedings of the 26th Pacific Asia Conference on Language, Information and Computation*, pages 171–180.
- Alina Andreevskaia and Sabine Bergler. 2006. Mining wordnet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. *In Proceedings of EACL 2006*.
- Bing Liu. Sentiment Analysis and Opinion Mining. 2012. *Synthesis Lectures on Human Language Technologies*.
- Bo Pang and Lillian Lee. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL)*, pp. 271–278.
- Bo Pang and Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. *Now Publishers Inc*.
- Carmen Banea, Rada Mihalcea, and Janyce Wiebe. 2010. Multilingual subjectivity: are more languages better. *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pp. 28–36.
- Carmen Banea, Rada Mihalcea, Janyce Wiebe and Samer Hassan. 2008. Multilingual subjectivity analysis using machine translation. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 127–135.
- Chenghua Lin, Yulan He and Richard Everson. 2011. Sentence Subjectivity Detection with Weakly-Supervised Learning. *Proceedings of the 5th International Joint Conference on Natural Language Processing*, pp. 1153–1161.
- Ellen Riloff and Janyce Wiebe. 2003. Learning Extraction Patterns for Subjective Expressions. *In: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 105–112.
- Ellen Riloff, Siddharth Patwardhan, and Janyce Wiebe. Feature subsumption for opinion analysis. 2006. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2006)*.
- Farah Benamara, Baptiste Chardon, Yannick Mathieu, and Vladimir Popescu. 2011. Towards Context-Based Subjectivity Analysis. *In Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP-2011)*.
- Gabriel Murray and Giuseppe Carenini. 2009. Predicting subjectivity in multimodal conversations. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1348–1357.
- Gregory Grefenstette, Yan Qu, David A. Evans, and James G. Shanahan. 2006. Validating the Coverage of Lexical Resources for Affect Analysis and Automatically Classifying New Words along Semantic Axes. *In: Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*, pp. 93–107.
- Hiroshi Kanayama and Tetsuya Nasukawa. 2006. Fully automatic lexicon expansion for domain-oriented sentiment analysis. *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 355–363.
- Janyce Wiebe and Ellen Riloff. 2005. Creating subjective and objective sentence classifiers from unannotated texts. *Computational Linguistics and Intelligent Text Processing, Springer*, pp. 486–497.
- Justin Martineau, and Tim Finin. 2009. Delta TFIDF: An Improved Feature Space for Sentiment Analysis. *In Proceedings of the 3rd AAAI International Conference on Weblogs and Social Media*.
- Kamil Aida-zade, Samir Rustamov, Elshan Mustafayev, and Nigar Aliyeva, 2012. Human-Computer Dialogue Understanding Hybrid System. *IEEE Xplore, International Symposium on Innovations in Intelligent Systems and Applications. Trabzon, Turkey*, pp. 1-5.
- Luciano Barbosa and Junlan Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. *In Proceedings of the International Conference on Computational Linguistics (COLING-2010)*.
- Muhammad Abdul-Mageed, Mona T. Diab, and Mohammed Korayem. 2011. Subjectivity and sentiment analysis of modern standard Arabic, *In Proceedings of the 49th Annual Meeting of*

- the Association for Computational Linguistics: short papers*, pages 587–591.
- Rada Mihalcea, Carmen Banea and Janyce Wiebe. 2007. Learning multilingual subjective language via cross-lingual projections. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 976–983.
- Robert Fuller. *Neural Fuzzy Systems*, 1995.
- Robert Remus. 2011. Improving Sentence-level Subjectivity Classification through Readability Measurement. *NODALIDA-2011 Conference Proceedings*, pp. 168–174.
- Samir Rustamov, Elshan Mustafayev, Mark Clements. 2013. Sentiment Analysis using Neuro-Fuzzy and Hidden Markov Models of Text. *IEEE Southeastcon 2013*, Jacksonville, Florida, USA.
- Smruthi Mukund and Rohini K. Srihari. 2010. A vector space model for subjectivity classification in Urdu aided by co-training. In *Proceedings of Coling 2010: Poster Volume*, pages 860–868.
- Soo-Min Kim and Eduard Hovy. 2005. Automatic Detection of Opinion Bearing Words and Sentences. In: *Companion Volume to the Proceedings of the International Joint Conference on Natural Language Processing*, pp. 61–66.
- Stephan Raaijmakers, Khiet Truong, and Theresa Wilson. 2008. Multimodal subjectivity analysis of multiparty conversation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 466–474.
- Taras Zagibalov and John Carroll. 2008. Unsupervised classification of sentiment and objectivity in Chinese text. In *Proceedings of International Joint Conference on Natural Language Processing (IJCNLP-2008)*, pp. 304–311.
- Theresa Wilson, Janyce Wiebe, Rebecca Hwa. 2004. Just How Mad Are You? Finding Strong and Weak Opinion Clauses. In: *Proceedings of the National Conference on Artificial Intelligence*, pp. 761–769.
- Yulan He. 2010. Bayesian Models for Sentence-Level Subjectivity Detection. *Technical Report KMI-10-02, June 2010*.
- Ziqiong Zhang, Qiang Ye, Rob Law, and Yijun Li. 2009. Automatic Detection of Subjective Sentences Based on Chinese Subjective Patterns. *Proceedings of 20th International Conference, MCDM-2009*, pp. 29-36.