Machine Learning with Dual Process Models

Bert Klauninger, Martin Unger and Horst Eidenberger
Institute for Interactive Media Systems, Vienna University of Technology, Vienna, Austria

Keywords: Dual Process Model, Similarity Measures, Combined Similarity Measures, SVM Kernels, Predicative Measurements, Quantitative Measurements.

Abstract: Similarity measurement processes are a core part of most machine learning algorithms. Traditional approaches focus on either taxonomic or thematic thinking. Psychological research suggests that a combination of both is needed to model human-like similarity perception adequately. Such a combination is called a Similarity Dual Process Model (DPM). This paper describes how to construct DPMs as a linear combination of existing measures of similarity and distance. We use generalisation functions to convert distance into similarity. DPMs are similar to kernel functions. Thus, they can be integrated into any machine learning algorithm that uses kernel functions.Clearly, not all DPMs that can be formulated work equally well. Therefore we test classification performance in a real-world task: the detection of pedestrians in images. We assume that DPMs are only viable if they yield better classifiers than their constituting parts. In our experiments, we found DPM kernels that matched the performance of conventional ones for our data set. Eventually, we provide a construction kit to build such kernels to encourage further experiments in other application domains of machine learning.

1 INTRODUCTION

Similarity measurement processes are a core part of most machine learning algorithms. Traditional approaches focus on either *taxonomic* ("A and B share properties x, y and z") or *thematic* ("A is similar to B by value N") thinking. Psychological research, e.g. (Wisniewski and Bassok, 1999), suggests that a combination of both is needed to adequately model human-like similarity perception.

Any model combining those aspects is called a *Similarity Dual Process Model*. The primary aim of our work is to provide an *implementation of the DPM idea* for computer vision. It should perform binary classification and be adoptable to carry out other machine learning tasks like, for example, cluster analysis, correlation and ranking.

The secondary aim of this work is to test DPMs in real-world experiments. The selected scenario should have intermediate applications, while still being simple enough to generate results within reasonable time. The question for the experimental results is: *Which DPM performs best?*

In the following section, we sketch the necessary background. In particular, we explain taxonomic and thematic thinking and the associated types of measures more deeply. Afterward, we turn to more technical aspects arising from the real-world task we selected: the detection of pedestrians in images. It has been chosen because of its interesting applications and because various well-known feature extraction algorithms already exist.

Section 4 describes our experimental setup and the results we obtained. In section 5, we start with a comparison to existing models and discuss the viability of DPMs. Next, we take a look at the effect of using different generalisation functions and measures. At this point, we are able to list the DPMs that performed best, thereby reaching our secondary goal. Eventually, section 6 gives a conclusion and mentions promising areas of further research.

2 BACKGROUND

2.1 Taxonomic vs. Thematic Thinking

Taxonomic thinking tries to identify common features and differences between objects. The more common features can be identified, the larger the similarity. Hence, taxonomic similarity assessment is associated with predicate based similarity measures ("counting"). Thematic thinking tries to find a theme that connects the objects. This theme is then used for comparison. This kind of reasoning is mostly associ-

Klauninger, B., Unger, M. and Eidenberger, H.
Machine Learning with Dual Process Models.
DOI: 10.5220/0005655901480153
In *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2016)*, pages 148-153
ISBN: 978-989-758-173-1

ated with metric distances ("measuring").

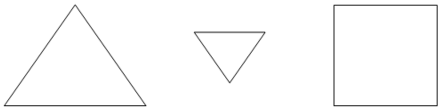


Figure 1: Taxonomic and Thematic Thinking (cf. (Eidenberger, 2012, p. 540)).

Figure 1 tries to develop an intuitive understanding with an example. The triangle on the left side is the reference. We compare it to the two stimuli next to it. If the focus lies on taxonomic thinking, the triangle in the center is more similar to the reference, because it also has three corners. If the focus lies on thematic thinking, the square on the right is less different from the reference, because it is similar in size.

To measure distance between two concepts, it is often constructive to find a commonality first. We call such concepts *alignable*. Highly alignable concepts tend to lead to taxonomic thinking, because there are enough common features to be able to make a comparison. Poorly alignable ones tend to lead to thematic thinking, because there are no common features to even work with taxonomic thinking.

Depending on whether we measure or count, different measures are used for taxonomic and thematic thinking. The dot product and the number of co-occurrences are typical choices for taxonomic thinking. The city block distance and the Hamming distance (features that are in one object, but not the other) are typical choices for thematic thinking.

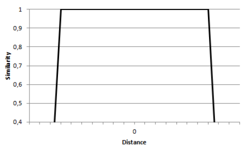
Table 1: Properties of Taxonomic and Thematic Thinking (cf. (Eidenberger, 2012, p. 537)).

Property	**Taxonomic**	**Thematic**
Stimuli	Separable	Integral
Concern	Similarity	Distance
Measurement	Dot Product	City Block Distance
Counting	Co-Occurrences	Hamming Distance

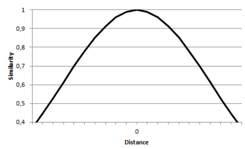
2.2 Generalisation

The relation between distances and similarities can be formalised with *generalisation functions*. The higher the difference between objects, the lower the probability of them belonging to the same class. Identical objects should belong to the same class with a probability of $p = 1$. From this point, probability should fall with similarity. How exactly it should fall, is still being disputed.

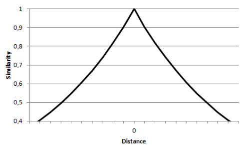
Generalisation functions allow us to convert distance into similarity. This is an important part of DPMs, because we need a way of combining similarity and distance. Note that we deal with negative distances with a symmetry assumption to simplify distance measurements.



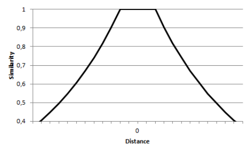
(a) Boxing.



(b) Gaussian.



(c) Shepard.



(d) Tenenbaum.

Figure 2: Different Generalisation Functions.

Figures 2a-2d show different generalisation functions. The Tenenbaum function is considered the state of the art.

2.3 The Dual Process Model

At first sight, the combination of both taxonomic and thematic similarity assessment seems an unnecessary complication. After all, both have been used on their own in machine learning - mainly in models created by computer scientists. Humans, however, do not always use one or the other approach when making similarity judgments. One experiment asked participants to rate the similarity of word pairs on a numeric scale, e.g. (milk, coffee), (milk, lemonade), (milk, cow) and (milk, horse). "As one would expect, similarity ratings for pairs which were highly alignable were reliably higher than for pairs which were poorly alignable. However, contrary to present accounts of similarity, ratings for pairs with preexisting thematic relations were higher than for pairs without preexisting thematic relation." (Wisniewski and Bassok, 1999, p. 216f)

$$s_{dpm} = \alpha\, s_{taxonomic} + (1-\alpha)\, g(d_{thematic}) \qquad (1)$$

Equation 1 shows a simple linear DPM (Eidenberger, 2012, p. 540), s_{dpm} being the total similarity score; $\alpha, 0 \leq \alpha \leq 1$ the importance of taxonomic thinking; $s_{taxonomic}$ the specific taxonomic measure (similarity); g the generalisation function; $d_{thematic}$ the specific thematic measure (distance).

The specific taxonomic measure, thematic measure and generalisation function to be plugged into the equation are the reader's decision. Note that, for the rest of this work, we deal with linear combinations of taxonomic and thematic thinking only.

3 IMPLEMENTATION

Our implementation reads training images and generates feature vectors from them. Our features are the *Histogram of Gradients* (HOG) (Dalal and Triggs, 2005), the MPEG-7 *Edge Histogram* (EHD) and the *Scalable Color Descriptor* (SCD) (Sikora, 2001). HOG divides images into parts and creates a histogram for each part. The histograms describe the gradient orientations of each part which can be calculated from the horizontal and vertical gradients. For EHD and SCD, the MPEG-7 reference implementation was used.

The aim of this work is not to come up with a new, high-performing image detection algorithm. Rather, the effect of using a DPM for measuring distance in existing algorithms is examined. We selected the HOG algorithm because it is representative for a line of research called gradient-based algorithms. Empirical research suggests to combine HOG with other feature extraction methods to obtain a stronger algorithm (Dollar et al., 2012, p. 10).

SCD and EHD from MPEG-7 provide this additional information to our implementation. Furthermore, turning these descriptor values into predicate based data is straightforward. SCD and EHD do not return predicates (i.e. zero/one values). To be able to work with predicates, the values returned by SCD and EHD are put into evenly sized bins by Algorithm 1. Note that the algorithm does something different than creating a histogram. The output is an array of values that are either zero or one.

Input: *binSize* size of one bin, *minT* smallest possible value, *binNum* number of bins to create, *values* array of values to be binned
Output: *binnedValues* array of binned values
for *i=0; i < values.size(); ++i* **do**
    for *b = 0; b < binNum; ++b* **do**
        val = values.at(i);
        if $val \geq minT + b \cdot binSize \;\&\; val < minT + (b+1) \cdot binSize$ **then**
            binnedValues[$i \cdot binNum + b$]=1;
        end
    end
end

Algorithm 1: Transformation of Quantitative Values into Predicates.

HOG, in the version we use, is not scale-invariant, while EHD and SCD are scale-invariant. Therefore we have to resize our images to the correct size. After this step, feature vectors are constructed in such a way that the first N elements should be treated as predicates, the remaining ones as distance measurements. We train a modified *Support Vector Machine* (SVM, (Joachims, 1998)) with the generated feature vectors using a DPM kernel. This results in a SVM model file containing the support vectors that create an optimal separation of the training data.

The pedestrian detection part extracts the feature vectors from the training set and uses the trained SVM model to classify them. We use one of the most straightforward methods for evaluating classifier quality: the correct classification rate. More advanced evaluation measures (precision, recall, ...) exist. Their analysis was out of scope for this paper.

DPMs stipulate the use of quantitative and predicate-based measures to represent taxonomic and thematic thinking. We do not mandate which type of measure to use for which type of thinking. We can combine a quantitative measure for taxonomic thinking with a predicate-based measure for thematic thinking or we can use only quantitative or only predicate-based measures.

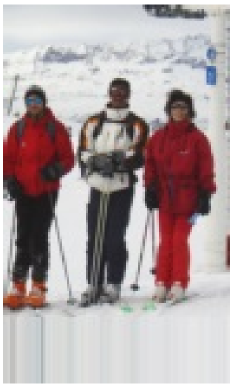
4 TEST ENVIRONMENT

We selected the INRIA dataset[1] with upright images of persons in everyday situations. The dataset is 970 MB large and contains thousands of images. Example images are shown in Figures 3a and 3b.

Training was performed with 140 positive and 160 negative samples, testing with 50 positive and 50 negative images. During SVM training, the number of allowed iterations without progress was restricted to 3000. We performed a manual classification into thematic and taxonomic measures. If there is a contrast (i.e. $x-y$, $\frac{x}{y}$, $\frac{a}{\cdot}$, $\frac{\cdot}{-a}$, $\frac{\cdot}{b}$ or $\frac{\cdot}{c}$), then a measure is thematic and belongs on the right-hand side of Equation 1. Otherwise, it is taxonomic and belongs on the left-hand side.

The importance of taxonomic thinking was set to $\alpha = \frac{1}{2}$ during all experiments. This means we simulate a person that values taxonomic thinking as much as thematic thinking. We ran pedestrian detection with the described dataset for all combinations of quantitative/predicate based measure/generalisation function. In order to restrict the search space, only predicate-based and quantitative measures were used that were part of a purely predicate-based or purely quantitative DPM that performed as good as the linear kernel. To be able to compare our DPMs to the cur-

[1] http://pascal.inrialpes.fr/data/human (last accessed 2015-02-24)



(a) Positive Example.

(b) Negative Example.

Figure 3: Examples From the Dataset.

rent state of the art, we also ran pedestrian detection with the described dataset for the linear, polynomial, sigmoid and radial kernels. Additionally, the thematic and taxonomic parts of each DPM were used on their own to classify the test data set.

Furthermore, we carried out five experiments with selected DPMs and a larger dataset. This dataset contained 2379 positive and 1231 negative training samples. Testing in this case was done with 900 samples. Experiments which did not terminate during training have been omitted from the result discussion.

5 EVALUATION

5.1 Viability of the Dual Process

Using a Dual Process Model adds complexity to any machine learning task. Is the overhead worth it? We compare the classification performance of each DPM with its single process models, i.e. the taxonomic part $m_{taxonomic}$ with the thematic part $g(m_{thematic})$. The aggregated results of this comparison are shown in Table 2. It was created by measuring the classification performance of each possible DPM and comparing it to the classification performances of the two single process models that belong to it.

We can see in the first two lines that taxonomic thinking on its own often performed better than the thematic thinking on its own. This can be seen as a clue that taxonomic thinking has a larger impact on

Table 2: Viability of DPMs.

Aspect	% of Cases
Taxonomic Thinking Better Than Thematic Thinking	71.12
Thematic Thinking Better Than Taxonomic Thinking	19.03
Single Process Model Better Than or Equal DPM (Not Viable)	73.77
Single Process Model Worse Than DPM (Viable)	26.23

image detection performance than thematic thinking. However, it is equally likely that this difference is caused by the way we combined taxonomic and thematic thinking in our model or by the algorithm selection for feature vector extraction.

The last two lines are more important. They tell us that DPMs are not necessarily better than single process models. In other words: Not every DPM created with Equation 1 makes sense. In about 74% of the search space, the classification performance has nothing to gain from the use of a DPM with fixed importance factor $\alpha = \frac{1}{2}$ and might even decrease. Therefore, when formulating a DPM, it is essential to verify that it improves performance for the task at hand.

Let us call this performance improvement *viability*. For the rest of the result discussion, we will exclude all DPMs that are not viable. It should be mentioned that using the Euclidean distance (a specific case of the Minkowski distance) often resulted in strong classification performance. However, because of our viability constraint, this measure does not appear often in the following results.

5.2 Comparison to Existing Models

Are DPMs better for image classification than the current state of the art? To answer this question, we compared them to linear, radial, polynomial and sigmoid kernels. Table 3 shows a summary of the classification performance for our pedestrian detection task.

Table 3: Comparison to the State of the Art.

Kernel	% Correct
Sigmoid	50
Radial	50
Linear	92
Polynomial	92
Baroni + Shepard (Normalisation)	96
Histogram + Shepard (Minkowski)	95
Tanimoto Index + Boxing	94
Russel & Rao-Minkowski	94

Experiments showed that sigmoid and radial kernels performed poorly, while the widely-used linear and polynomial kernels performed well. Many tested

DPMs made less than 90% correct test data classifications. However, about 9% of all DPMs performed that well or better.

5.3 Generalisation

All DPMs were tested with different generalisation functions. We group all DPMs with a classification performance of at least 90% by their generalisation function. The results can be seen in Table 4.

Table 4: Percentage of High-performing DPMs per Generalisation Function.

Generalisation	% of good DPMs
Shepard	51,0
None	23,5
Gaussian	23,5
Boxing	2,0

The Shepard generalisation function was most often part of high-performing DPMs. Surprisingly, not using any generalisation function proved as successful as using the Gaussian generalisation function. The Boxing function did not work well, but we found that its performance increases if additional iterations were allowed during SVM training.

The data show that if a combination of taxonomic and thematic measure is successful with one generalisation function, it tends to be successful with other generalisation functions, too. The probability of arriving at a high-performing DPM is highest when using the Shepard generalisation function. However, good classification performances could be obtained with most generalisation functions, as long as fitting thematic and taxonomic measures were chosen. Based on the experiments, selecting the taxonomic and thematic measure seems to have a much larger impact on the classification performance of DPMs.

5.4 Quantitative and Predicate-based Measures

As already discussed, quantitative measures operate on real-valued feature vectors, while predicate-based measures operate on 0/1 values. To keep the number of experiments manageable, we first had to test all purely quantitative and all purely predicate-based DPMs. Only the most promising measures of these experiments where tested in combination. This approach is inspired by genetic algorithms.

The experiments indicated that if quantitative measures are used exclusively, their performance is slightly better than the exclusive use of predicate-based measures. Table 5 states the classification performance of the best DPMs that use only predicate-based measures.

Table 5: Classification Performance of Predicate-based DPMs (Shepard Generalisation used in all instances).

Taxonomic	Thematic	%
Sorgenfrei	Batagelj & Bren	90
Hawkins & Dotson	Variance Dissimilarity	90
Baroni-Urbani & Buser	Baulieu Variant 2	90
Coeff. of Arith. Means	Baulieu Variant 2	90
Proportion of Overlap	Baulieu Variant 2	90

Like before, 90% seems to appear more often than it should. Again, this is explained by the difficult test images that lead to the same errors for all shown DPMs. Hence, our predicate-based feature vector extraction is not discriminative enough.

Table 6 shows the best DPMs that use only quantitative measures. Their correct classification rate is always a little bit higher than the rate of their predicate-based counterparts.

Table 6: Classification Performance of Quantitative DPMs.

Taxonomic	Gen.	Thematic	%
Histogram Intersection	Shepard	Minkowski Dist., Meehl Index	95
Histogram Intersection	Shepard	Kullback/Leibler, Jeffrey Divergence	95
Histogram Intersection	Shepard	Exp. Divergence, Normalisation	95
Histogram Intersection	Shepard	Kagan Divergence, Mahalanobis Dist.	95
Tanimoto Index	Shepard	Minkowski Dist.	94
Modified Dot Product	Gauss	Minkowski Dist., Mahalanobis Dist.	93
Modified Dot Product	Shepard	Mahalanobis Dist.	93
Cosine Measure	Gauss	Minkowski Dist., Mahalanobis Dist.	93
Cosine Measure	Shepard	Mahalanobis Dist.	93
Tanimoto Index	Shepard	Normalisation, Mahalanobis Dist.	92

Until now, our DPMs used either quantitative or predicate-based measures only - but these two types of measures can be mixed. Table 7 summarises the classification performance of the best mixed DPMs.

Note that some mixed DPMs appear in Table 7 that were not part of the best purely quantitative DPMs or purely predicate-based DPMs. The reason for this is that DPMs that performed well, but were not viable, were also permitted to take part in the mixed test round. However, all of the mixed DPMs are still required to be viable. We can see that mixed DPMs work as well as quantitative or predicate-based DPMs. This is an encouraging result, because it allows us to select our DPM parts based on the feature vector type at hand.

Table 7: Top 10 Classification Performance of Mixed DPMs.

Taxonomic	Gen.	Thematic	%
Baroni-Urbani & Buser	Shepard	Normalisation	96
Sorgenfrei	Shepard	Normalisation	95
Coeff. of Arith. Means	Shepard	Normalisation	95
Russel & Rao	Shepard	Exponential Div.	94
Russel & Rao	None	Minkowski Dist.	94
Russel & Rao	None	Exponential Div.	94
Tanimoto Index	Boxing, Gaussian	Compl. of Hamming Dist.	94
Tanimoto Index	Shepard	Compl. of Hamming Dist.	94
Correlation Coefficient	Shepard, None	Baulieu Var. 2	94
Correlation Coefficient	Shepard, None	Batagelj & Bren	94

Against intuition, mixing quantitative measures with predicate-based measures (that performed slightly weaker in general), still often lead to improved classification performance. This is further empirical evidence in support of DPMs.

6 CONCLUSIONS AND FUTURE WORK

We implemented pedestrian detection in images to be able to test DPMs in a real world task. Any DPM is a combination of two measures. Obviously, if using just one measure performs as well or better than using two measures, we do not deal with a viable DPM. Only 14% of our DPMs were found to be viable. DPMs can be formulated with quantitative measures (i.e. real values), predicate-based measures (i.e. countable or $0/1$ values) and with a mix of both types of measures. We did not find conclusive evidence that a certain measure type (e.g. measuring taxonomic and thematic thinking with quantitative measures) works better than any other type.

We discovered DPMs that performed as good as or better than the existing linear and polynomial kernels. However, it has to be mentioned that this is not a mandatory proof that DPMs outperform the current state of the art. Conclusive evidence would have to carry out statistical testing to be able to state significance levels of classification performances. For this, every single specific DPM has to be tested many times. To make this possible, runtime has to be improved.

Future work could either use algorithms that yield good classification results with much smaller feature vectors or focus only on a few possible DPMs. To support this, we provided a construction kit for well-performing DPMs. Another interesting direction for further research are DPMs in other domains, for example audio and text retrieval or non-multimedia problems like recommender systems and computational finance. Because DPMs are kernel functions, they can be readily used with algorithms other than SVMs like Gaussian processes, ridge regression, spectral clustering and many more.

REFERENCES

Dalal, N. and Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. In *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE.

Dollar, P., Wojek, C., Schiele, B., and Perona, P. (2012). Pedestrian Detection: An Evaluation of the State of the Art. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(4):743–761.

Eidenberger, H. (2012). *Handbook of Multimedia Information Retrieval*. BoD–Books on Demand.

Joachims, T. (1998). Making Large-Scale SVM Learning Practical. LS8-Report 24, Universität Dortmund, LS VIII-Report.

Sikora, T. (2001). The MPEG-7 Visual Standard for Content Description-An Overview. *Circuits and Systems for Video Technology, IEEE Transactions on*, 11(6):696–702.

Wisniewski, E. J. and Bassok, M. (1999). What Makes a Man Similar to a Tie? Stimulus Compatibility With Comparison and Integration. *Cognitive Psychology*, 39(3):208–238.