A Hybrid CBR-ANN Approach to the Appraisal of Internet Domain Names

Sebastian Dieterle and Ralph Bergmann

Business Information Systems II
University of Trier
54286 Trier, Germany
sebastiandieterle@web.de, bergmann@uni-trier.de
www.wi2.uni-trier.de

Abstract. Good domain names have become rare and trading with premium domain names has developed into a profitable business. Domain appraisals are required for many different reasons, e.g., in connection with a loan on a domain name. The aim of this paper is to analyze various methods for estimating prices for domain names. The criteria for this are predictive accuracy, traceability and speed of the appraisal. First, the scientific relevance of the topic is demonstrated based on intensive literature and Internet research. Several approaches based on artificial neural networks (ANNs) and case-based reasoning (CBR) are developed for estimating domain name prices. In addition, hybrid appraisal approaches are introduced that are built up on CBR and which use ANN for improved adaptation and similarity determination. The approaches are evaluated in several configurations using a training set of 4,231 actual domain transactions, which demonstrates their high usefulness.

Keywords: Internet Domain Names, Artificial Neural Networks, Hybrid Application, Appraisal, Similarity, Adaptation.

1 Introduction

Domain names are often seen as the land, and web sites as the buildings, of the virtual world. Due to their uniqueness, premium domain names achieve high market prices. There are many reasons why the fair market value of domain names must be appraised. The manual appraisal of domain names is subjective, time-consuming and expensive [1]. Existing approaches to the application of case-based reasoning (CBR) in the appraisal of assets [6–9] show three fundamental weaknesses: First, the rules for price adjustment often rely on the experience of experts and are therefore not empirically justified. Second, optimized weights are not used in determining similarity, or, third, optimized weights are only applicable to numeric attributes. Knowledge-intensive similarity measures are not used in appraisal methods based on locally weighted regression (LWR). The realization of an appraisal by means of an artificial neural network (ANN) [11, 12] is not traceable for users due to its black box character and is thus not allowed for an official legal appraisal [13]. In order to address these weaknesses, we propose hybrid approaches that combine CBR and ANN in different ways. We apply the traditional CBR approach to retrieve recently sold, similar domain names from a case base, and adapt the sales price of the cases with respect to the relevant differences to the query, i.e., the domain name to be appraised [9]. In this process, the ANN is used for two purposes: a) the weights for determining similarity are learned by an ANN, and b) the parameters of the adaptation rules for adjusting the price are also determined by an ANN. In addition, a different combination of CBR and ANN is proposed in which CBR is applied to pre-select cases to train an ANN being used for an appraisal based on LWR.

In Section 2, we introduce the related work. The basic approaches to case-based and neural domain appraisal are described in Section 3. In Section 4, three hybrid approaches are presented. An empirical evaluation in Section 5 based on 4,231 cases of domain transactions tests various hypotheses by experiments. The last section concludes with a summary and an outlook on future work.

2 Related Work

An asset can be evaluated in three different ways: based on acquisition costs, on income, or on market price [2]. The value of an asset (such as real estate) is determined by an appraiser finding recently sold real estate with similar characteristics in the neighborhood. These prices must be adjusted to increases and deductions, since no two properties can be compared exactly [3]. The 3Cs appraisal model from `GreatDomains.com` was the first to describe factors for domain appraisal. By means of a matrix, the criteria of *characters* (number of characters), *commerce* (commercial potential) and *.com* (value relevance of the TLD) determine the value of a domain. Multiple linear regression analyses were generally used. In the hedonic regression according to Phillips [4], the time factor is taken into account by pre-processing the data with the Morgan Stanley Internet Index. The distinguishing feature of Jindra's regression analysis [5] is that more than one regression model is calculated. Instead, the data are split up into four clusters and a regression equation is determined for each cluster.

The first use of CBR for the appraisal of real estate was published by Gonzalez & Laureano-Oritz [6]. Compared to regression analysis, this technique more closely resembles the way real estate appraisers work and is easier for users to understand and trace. The prices of the most similar previous cases are adjusted in accordance with heuristic rules and a weighted average value is calculated. The case-based appraisal of rental prices in the retail trade was studied by O'Roarty et al. [7]. They transform the rental prices from different years by means of a rental price index to a standardized level and leave any further price adjustment to the user. McSherry [8] presents a domain-independent adaptation heuristic based on the assumption of an additive valuation function and the existence of certain specific case pairs. The case-based appraisal of domain names by Dieterle & Bergmann [9] uses knowledge-intensive similarity functions. The adjustment of the prices is based, among other things, on the Internet Domain Name Index

(IDNX). The case-based appraisal approaches often use heuristic domain knowledge for adjusting prices and determining similarity. Furthermore, price indices consider only the time aspect; standardized regression coefficients are restricted to numeric attributes. In this paper, we shall introduce approaches for the learning of adaptation rules that do not require any specific case pairs and for the learning of weights for the aggregation of arbitrary local similarity measures.

ANN constitute a model which is inspired by the nerve activities of the brain and which allows, among other things, the approximation of linear and non-linear functions [10]. In the application field of the appraisal of real estate, linear regression is compared to a multi-layer perceptron (MLP) trained with backpropagation. Whereas Rossini [11] came to the conclusion that linear regression produces lower errors, Peterson & Flanagan [12] came to the opposite conclusion on a significantly larger data set. In comparison with CBR, however, both methods are less easy for the user to understand and trace, since the appraisal is based only implicitly on previous sales transactions. Because of the black box nature of ANNs, people are unable to trace the appraisal process and it must not be used as the basis for an official legal appraisal [13].

Hybrid methods make use of the existence of different strengths and weaknesses of the individual methods. On the basis of a linear regression analysis, Rossini [14] calculates additive adjustment rules, to use these by means of the technique of the k nearest neighbors to adjust previous transactions in real estate appraisal. Al-Akhras [15] uses an evolutionary genetic algorithm in order to determine the best topology for an MLP for real estate appraisal, and then trains this MLP with backpropagation. Jalali & Leake propose a hybrid approach to estimate car and house prices [16]. Here, the squared error weighted by an exotic distance function over the k nearest neighbors to a query case is minimized to determine a regression model and thus to appraise a query. One difference between the mentioned hybrid approaches and the approach we present in this paper is the fact that we use a knowledge-intensive similarity measure, so that the determined cases are also similar from a semantic point of view.

3 Case-Based and Neural Network Approaches

3.1 CBR for Domain Appraisal

In our previous work [9], we presented an approach to the case-based appraisal of domain names based on a case base with previous sales transactions. Case attributes are selected such that they allow an appropriate determination of similarity between two domain names. The sales price attribute is the solution attribute stored in each case. The overall CBR approach applied is shown in the top box of Fig. 1. In order to estimate the value of a domain name, it is entered as a query. In the *recall* step, the relevant features of the corresponding domain name are derived and further data are extracted from the Internet, leading to an enriched query (see also the left column of Table 1). In the *retrieve* step, a number of k similar cases are determined from a case base. For this purpose, a knowledge-intensive similarity measure based on the derived features is used.

A weighted average of the local similarity values is used to compute the global similarity. The weights are determined in a heuristic process by experimenting with several variants. In the *reuse* step, the solution attribute (i.e., the sales price) is adapted for each of the k most similar cases. The price index IDNX is used as heuristic expert knowledge for adaptation. In the *remove* step outliers are excluded from further calculations and in the *reckon* step a weighted average of the remaining values is calculated. The weights used for this purpose are derived from the similarity between case and query. This weighted average value represents the estimated value for the query.

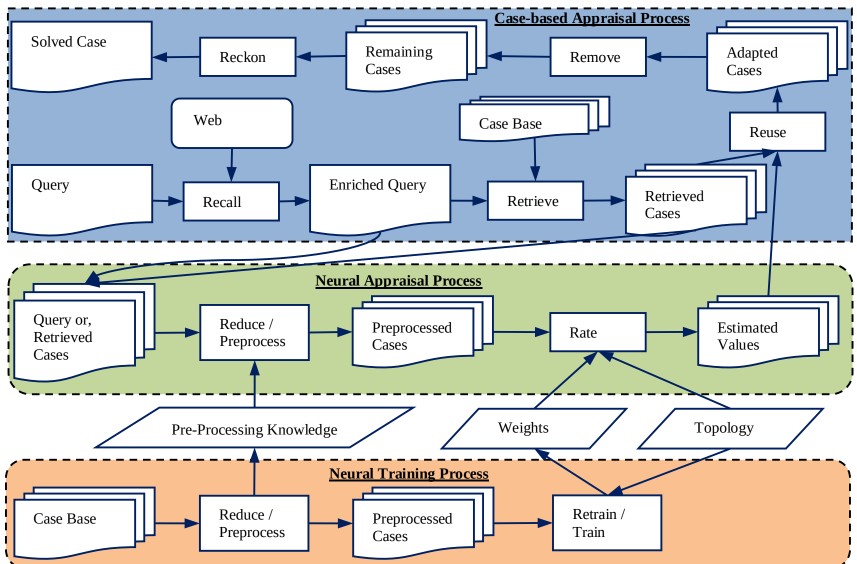


Fig. 1. Hybrid CBR/ANN approach for domain appraisal (this may also serve as an illustration for individual methods)

3.2 ANN for Domain Appraisal

Our overall neural approach to the appraisal of domain names [17] consists of a training process and an appraisal process (see the two bottom boxes in Fig. 1). In the *training process*, a *reduce* step transforms the case base from its original symbolic representation (Section 3.3) into a numeric vector representation suitable for an ANN. This transformation performs a stemming (Section 3.4), the derivation of binary attributes, and a normalization (Section 3.5). Thereby, pre-processing knowledge is generated and stored (e.g., attribute specific normalization parameters or correspondences between binary attributes and stemmed words). The pre-processed training set is used to *retrain* or train a given network structure of an ANN, such as an (adaptive) linear neuron (Adaline) or a

MLP with supervised learning, such as resilient propagation (Section 3.6). This method should be able to replicate the prices of the training examples as accurately as possible and at the same time be able to generalize to new problems. The network weights learned during the training are saved. The *appraisal process* describes the procedure for estimating the price of a query domain name. For this purpose, the query is enriched with the data *recalled* from the web (as in the case-based approach) leading to the enriched case representation of the query. In the *reduce* step, this query representation is transformed into the vector format using the same processing steps as in the training phase. Finally, in the *rate* step, the pre-processed query is presented as input to the trained network and the computed network output produces the estimated price for the query domain.

3.3 Case Representation

For the neural and case-based approach, we use two different case/data representations; however, in the hybrid models (Section 4) we combine the two models. The case representation follows a structural CBR approach and contains multi-valued, taxonomic, numeric and textual attributes, in order to enable knowledge-intensive similarity measures [9]. However, the neural data model consists of a numeric vector with a large number of binary and numeric attributes. In Table 1, the two models are compared, showing the case of "winterreise.de" (German for "winter journey").

The attribute "sales price", which describes at what price the domain name was traded, is the only solution attribute and output neuron. Differently, the transaction year is taken into account in the neural data model, with a range of seven binary attributes for each year, in order to take into account fluctuations in price level over time. The length of a domain indicates the number of characters contained in the second-level domain (SLD). This feature is considered to be a key criterion for the domain value, since short domain names can be more easily remembered and typed [4]. The case-based approach contains the multi-valued attribute "categories" (in the Open Directory Project (ODP)), which allows a hierarchical classification of all websites worldwide and the use of taxonomic similarity measures.

The attribute "word components" differs between the two models: Whereas in the case-based model the word components "winter" and "reise" are the values of this multi-valued attribute and allow the use of textual similarity measures, in the neural data model only binary attributes are used for frequent word stems.

Additional attributes describe how many results there are for a term in a search engine, whether the term contains hyphens, special characters or numbers and the age of the domain. Furthermore, consideration is given to how often a term is searched for in Google worldwide and in the region related to the top-level domain (TLD), the average cost per click (CPC), and the number of clicks.

Our approaches can also be applied to other data sets and application domains (e.g. the appraisal of businesses, art, cars, etc.). For this purpose, a different case description (and appropriate local similarity measures) must be defined.

Table 1. Description of the training example "winterreise.de" from 2007

Attribute	Example (CBR)	Example (ANN)
Domain name	`winterreise.de`	
SLD	`winterreise`	
TLD	`de`	
Transaction year /	2007	
2006, 2007, . . . , 2011, 2012		0; 1; 0; 0; 0; 0; 0
Length	11	11
Single- / 2- / 3-letter domain		0; 0; 0
Number of words	1	1
Number of word components	2	2
Categories (in the ODP)	World: German: Recreation: Travel: Travelogues; World: German: Sports: Winter Sports: Skiing: Journeys	
Words	Winterreise	
Word components /	Winter; Reise	
24, angebo, . . . , reis, . . . , www		0; 0; . . . ; 1; . . . ; 0
Search results	158,000	158,000
Contains hyphen	false	0
Number of hyphens	0	0
Contains special characters	false	0
Contains numbers	false	0
Domain age	1999.08333333	1999.08333333
Global monthly searches	5,400	5,400
Local monthly searches	1,600	1,600
Avg. CPC in €	0.81	0.81
Daily clicks	4.52	4.52
Daily cost in €	3.64	3.64
Sales price in €	10,000	10,000

3.4 Reduction Step and Stemming

In the reduction step, the training examples and the query are transformed into a vector form applicable to the ANN. For this purpose, binary attributes are introduced, on the one hand for the transaction year, and on the other hand for frequent word stems. Pre-processing knowledge is saved in regards to which binary attributes exist and occur in at least ten cases. Ten is a usual rule of thumb (see [18]) and represents a compromise between the regression model having a good capability to generalize and taking into account the maximum amount of information possible. First of all, the SLD was decomposed into the morphemes contained within it. For this purpose, the word is cut into two parts ($bookworm, b$ookworm, bo$okworm, etc.) and in each case the number of hits in the search engine is determined. Assuming that it exceeds a predefined threshold, the second most frequent spelling contains the SLD, separated into morphemes. To reduce the morphemes to their stem form (stemming), on one hand the approach in accordance with Caumanns [19] and on the other hand an algorithm defined by a snowball script[1] is used for German words running one after the other. In the first step, the process replaces particular strings in the words to be stemmed, e.g., replacing mutated vowels by the corresponding vowel. In the second step, suffixes are pruned away using a set of rules.

3.5 Normalization by Logarithmizing

Besides the semantic attributes, the numeric attributes must also be transformed into a form applicable to the ANN. This concerns, on the one hand, the logarithmic transformation of the data and, on the other hand, the compression of the data into a specific values range. The log of all the input values and of the output value is calculated and in addition an attribute-specific constant is included. This frequently used pre-processing step smoothes rapidly increasing values ranges.

3.6 Training an ANN with Resilient Propagation

Due to our data model, the ANN has 112 input neurons and one output neuron for the estimated price. The linear neuron (Adaline) and the MLP are considered as a topology. The MLP contains inner layers with the hyperbolic tangent function as an activation function and allows for the approximation of non-linear functions (non-linear regression). The linear neuron has no inner layers and allows a linear regression analysis. The logarithmic pre-processing of the attribute values results in a multiplicative value relationship between input and output values.

A variant of resilient propagation [21], iRPROP+ [20], is used as a supervised learning method to train the neural network. Here, the harmful influence of the partial derivation (i.e., the risk of a too great weight adjustment by a gradient, which is too steep at particular positions) is avoided by taking into account only the algebraic sign of the gradient. For each weight, this iterative method has its own weight-specific update value, which changes during the learning process.

4 Hybrid Approaches to Domain Appraisal

We now describe three variants for the integration of CBR and ANN. A common characteristic of our hybrid approaches is that an extended query description is generated and that similar cases are retrieved and reused in some way. Moreover, a neural network is trained and used for different purposes.

4.1 Hybrid ANN Adaptation

The first hybrid method applies the ANN using the Adaline topology to determine multipliers for the adaptation of the solution attribute, i.e., the sales prices

[1] `http://snowball.tartarus.org/algorithms/german2/stemmer.html`

of the retrieved cases. Figure 1 as a whole – and in particular the links between the case-based and the neural appraisal box – show the links between both methods. This hybrid appraisal process begins when a new query domain q is entered. As in the pure CBR approach, the *recall* and the *retrieve* step are performed and the k most similar cases for the query are determined using a knowledge-intensive similarity measure. The subsequent *reuse* phase is supported by the ANN. As described in Section 3, the neural appraisal process is capable of rating domain names on a stand-alone basis. The query and the k most similar cases are now assessed by means of the neural appraisal process. Thus, the *rate* step generates as output the estimated value for them. $v_n(q)$ is the estimated value for the query and $v_n(c)$ is the estimated value for a retrieved case c. These estimated values are now used in the *reuse* phase of the case-based appraisal process to adjust the price attribute of the k most similar cases according to the query. The adapted sales price of a case c, called ${c'}_p$, results from the sales price obtained from a case c_p multiplied by the ratio between the neural estimated value of query q and case c:

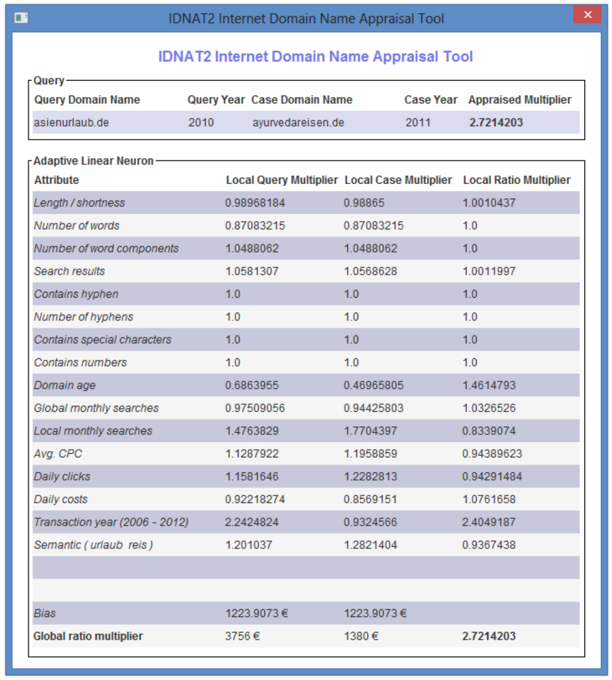
$$c'_p := c_p \cdot \frac{v_n(q)}{v_n(c)} \tag{1}$$

Due to the logarithmic pre-processing of the input and output data in the linear neuron, the estimated values $v_n(q)$ and $v_n(c)$ can be split into a basic value (bias) and local multipliers for every attribute (cf. [17]). This applies to the quotient of the above equation, leading to the following adaptation formula:

$$c'_p := c_p \cdot \prod_{a \in A} \underbrace{\frac{(i_a(q_a))^{w_a}}{(i_a(c_a))^{w_a}}}_{local\ multiplier} \tag{2}$$

The actual sales price of a case is thus adapted by a number of local multipliers, one for each attribute $a \in A$. A local multiplier results from the quotient of an attribute-specific value for query and case. Here, $i_a(.)$ is the normalization function applied to the case/query value q_a/c_a of attribute a. w_a is the weight of attribute a that results from the training of the Adaline. A valuation of the two domain names "asienurlaub.de" (query) and "ayuvedareisen.de" (case) is illustrated in Fig. 2. For the case and the query, the second and third column, respectively, show the multipliers per attribute. The last line shows the resulting price estimated by the output neuron.

The following steps proceed according to the CBR appraisal approach, i.e., outliers are *removed* and the weighted average value of the remaining similar cases is determined (*reckon*). Fig. 3 shows results of the hybrid appraisal of the query "asienurlaub.de". The eleven most similar cases are determined, the relevant differences between case and query are adjusted by means of multipliers, and a weighted average price is calculated.



IDNAT2 Internet Domain Name Appraisal Tool

IDNAT2 Internet Domain Name Appraisal Tool

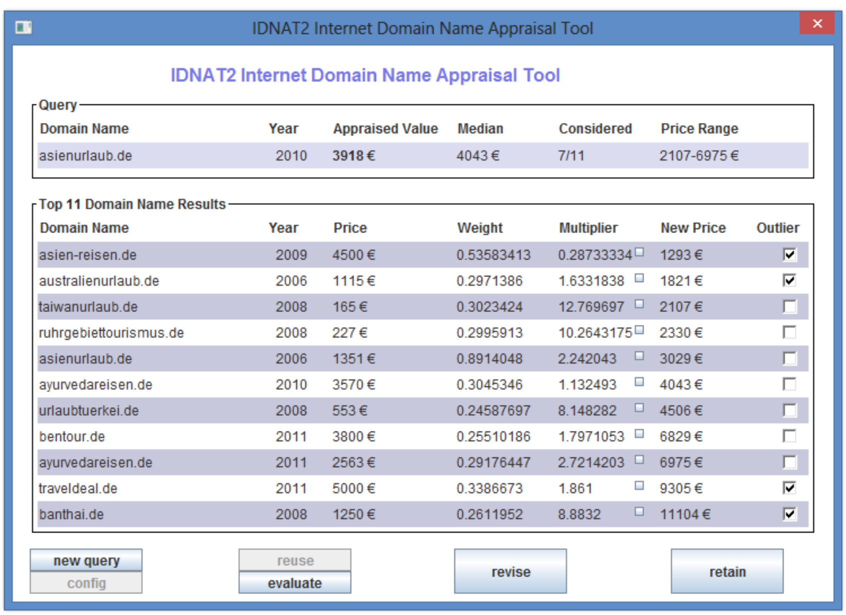
Query

Query Domain Name	Query Year	Case Domain Name	Case Year	Appraised Multiplier
asienurlaub.de	2010	ayurvedareisen.de	2011	**2.7214203**

Adaptive Linear Neuron

Attribute	Local Query Multiplier	Local Case Multiplier	Local Ratio Multiplier
Length / shortness	0.98968184	0.98865	1.0010437
Number of words	0.87083215	0.87083215	1.0
Number of word components	1.0488062	1.0488062	1.0
Search results	1.0581307	1.0568628	1.0011997
Contains hyphen	1.0	1.0	1.0
Number of hyphens	1.0	1.0	1.0
Contains special characters	1.0	1.0	1.0
Contains numbers	1.0	1.0	1.0
Domain age	0.6863955	0.46965805	1.4614793
Global monthly searches	0.97509056	0.94425803	1.0326526
Local monthly searches	1.4763829	1.7704397	0.8339074
Avg. CPC	1.1287922	1.1958859	0.94389623
Daily clicks	1.1581646	1.2282813	0.94291484
Daily costs	0.92218274	0.8569151	1.0761658
Transaction year (2006 - 2012)	2.2424824	0.9324566	2.4049187
Semantic (urlaub reis)	1.201037	1.2821404	0.9367438
Bias	1223.9073 €	1223.9073 €	
Global ratio multiplier	3756 €	1380 €	**2.7214203**

Fig. 2. Determination of multipliers for a query and a similar case



IDNAT2 Internet Domain Name Appraisal Tool

IDNAT2 Internet Domain Name Appraisal Tool

Query

Domain Name	Year	Appraised Value	Median	Considered	Price Range
asienurlaub.de	2010	**3918 €**	4043 €	7/11	2107-6975 €

Top 11 Domain Name Results

Domain Name	Year	Price	Weight	Multiplier	New Price	Outlier
asien-reisen.de	2009	4500 €	0.53583413	0.28733334	1293 €	☑
australienurlaub.de	2006	1115 €	0.2971386	1.6331838	1821 €	☑
taiwanurlaub.de	2008	165 €	0.3023424	12.769697	2107 €	☐
ruhrgebiettourismus.de	2008	227 €	0.2995913	10.2643175	2330 €	☐
asienurlaub.de	2006	1351 €	0.8914048	2.242043	3029 €	☐
ayurvedareisen.de	2010	3570 €	0.3045346	1.132493	4043 €	☐
urlaubtuerkei.de	2008	553 €	0.24587697	8.148282	4506 €	☐
bentour.de	2011	3800 €	0.25510186	1.7971053	6829 €	☐
ayurvedareisen.de	2011	2563 €	0.29176447	2.7214203	6975 €	☐
traveldeal.de	2011	5000 €	0.3386673	1.861	9305 €	☑
banthai.de	2008	1250 €	0.2611952	8.8832	11104 €	☑

new query | config | reuse | evaluate | revise | retain

Fig. 3. GUI of the hybrid price adjustment of domain names

4.2 Hybrid ANN Similarity

The purpose of the second hybrid approach is to improve the assessment of similarity by the optimization of the weights used in the weighted average aggregation of the local similarity values into the global similarity. Please note that we use various knowledge-intensive local similarity measures, such as numeric, taxonomic, and textual similarity measures (see 3.1 and [9]). The hybrid approach is illustrated in Fig. 4, showing the neural training and the case-based appraisal process.

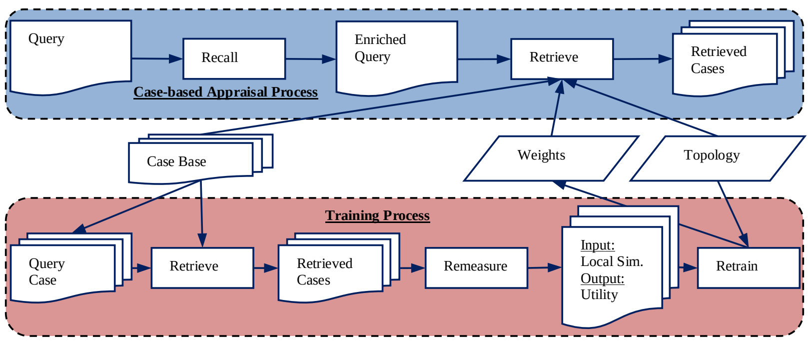


Fig. 4. Optimization of weights by a neural network

The training process makes use of the existing case base. Iteratively, each case in the case base is used as a query q and the k most similar cases to it are determined (*retrieve* step) by means of the non-optimized (initial) similarity measure. The retrieved query-case pairs (q, c) are used to derive training data for the ANN (which is different to the ANN of the first hybrid approach). To ensure a sufficient amount of training data, a relatively large value for k is chosen (e.g., $k = 50$). In the *remeasure* step, for each case pair the *utility* of the case for the query is determined, i.e., how well the sales price c_p of the case c predicts the sales price q_p of the query q.

$$utility(q, c) := \frac{min(q_p, c_p)}{max(q_p, c_p)} \tag{3}$$

This utility value measures how similar (and therefore how useful) two solutions are to each other, i.e., it is a kind of solution similarity. For instance, a utility of 0.5 results if the price of the case is double or half the price of the query. The subsequent *retrain* step uses the linear neuron as a topology and resilient propagation as the learning method. Each query-case pair provides a training sample. In particular, the vector of the local similarity values between case and query attribute values represent the net input. The utility value represents the desired net output.

Due to the optimization process which the ANN performs during learning, the attribute weights are adjusted as follows: given a query-case pair with a high utility value, the attributes with a high local similarity will get a high weight. Conversely, attributes with a low similarity will receive a low weight. Thus, the *utility* value and the weighted sum of the local similarity measures should approximate one another. The network thus minimizes the difference between these two values across all query-case pairs. The purpose of this is, on the one hand, to weight cases that have achieved similar sales prices in such a way that they achieve the highest possible similarity to one another and, on the other hand, to weight cases that show different sales prices in such a way that they achieve the lowest possible similarity to one another. This should ensure that the most similar cases are also the cases which have, as far as possible, achieved the most similar prices.

Similarly, an MLP trained with utility values can also be used to transform the local similarity measures into a global similarity measure by forward propagation. However, the resulting similarity assessment is less transparent, as the forward propagation cannot be traced.

4.3 Hybrid LWR based on CBR

The last hybrid approach uses case-based retrieval, i.e., a pre-selection by means of CBR, as a pre-processing step for an ANN. This results in a form of LWR, in which cases which lie further away from the query have no influence on the determined regression model. The approach consists of three phases, as illustrated in Fig. 5: a case-based retrieval process, a neural appraisal process, and a neural training process.

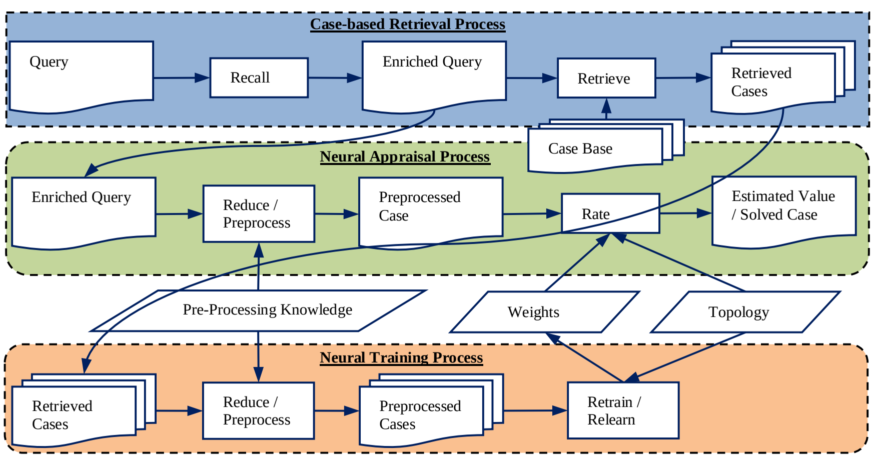


Fig. 5. Case-based pre-selection for a neural network

In the case-based *retrieval* process, the k most similar cases are determined for the query by means of a knowledge-intensive similarity measure. For this purpose, a relatively large value for k is selected (e.g., 20 % of the case base), in order to be able to determine a regression model capable of generalization.

The reuse step is now implemented in the form of the neural appraisal and training process. In the *reduce* step, the query on the one hand and the k most similar cases on the other hand are transformed from the case-based format into the numeric vector format, as described before. In the *retrain* phase the ANN is trained for each query with the preselected training set, i.e., the determined k most similar cases. This repeated training for each individual query is a major difference to the two previous hybrid approaches, in which the training is executed only once. This has the consequence that the linear neuron is especially optimized for the local (and semantic) neighborhood of the query. The training processes minimizes the following error function:

$$E := \sum_{c \in Ret(q)} (f(c) - c_p)^2 \tag{4}$$

Thereby, $Ret(q)$ represents the k most similar cases retrieved for the query q, c_p is the actual sales price of case c, and $f(c)$ is the value predicted by the ANN for c. The LWR only requires a simple topology, and often even a linear model is sufficient. In the *rate* step which follows, the pre-processed query is presented to the retrained ANN as input and thus an estimated value for the query is calculated by forward propagation. A fundamental difference to the ordinary LWR is that, in this approach, semantically similar cases are retrieved by a knowledge-intensive similarity measure.

5 Empirical Evaluation

A prototype of the proposed approach has been implemented, called Internet Domain Name Appraisal Tool Version 2 (IDNAT2). It is implemented in JAVA using various program libraries, such as jCOLIBRI2 for CBR, Encog3 for ANN, Apache Commons for statistical functions, Apache Lucene for stemming functions, and Jsoup for HTML decoding. In order to perform an experimental evaluation, a case base consisting of 4,231 cases describing domain sales transactions with the TLD .de was extracted from the Internet. For this purpose, the domain transaction list of the United-Domains AG was used, since it is easily accessible and relatively comprehensive with over 1,000 .de entries. Furthermore, the world's largest public list from Namebio - with over 3,000 relevant entries - was also used.

We have considered three quantitative criteria and one qualitative criterion in order to measure the quality of the appraisal. The standard evaluation criterion of the predictive accuracy of a regression analysis is the squared correlation coefficient R^2 (in this study, between the predicted value calculated by IDNAT2 and the actual sales price). However, there are some very highly priced values among the domain prices, which is why we use logarithmic values ($log(v+10)$) to keep the positive or negative influence of single values on the R^2 small. Moreover, we have measured the time required to solve an appraisal query and the time required to train the ANN. As a qualitative criterion, we have considered how traceable - and thus how reliable - the solution is. As a validation design for all the experiments, an n-fold cross validation was used with four test sets performed on an ASUS desktop PC (CM6870 Series).

In the context of the evaluation, the following three hypotheses were tested:
H1: The hybrid ANN adaptation achieves higher predictive accuracy than CBR without adaptation.
H2: The hybrid ANN similarity (with optimized weights) achieves higher predictive accuracy than CBR with heuristic, manually assigned weights.
H3: The hybrid LWR based on CBR achieves higher predictive accuracy than a pure neural network (Adaline or MLP).

Table 2. Comparison of the appraisal approaches on the test set

Approach	R^2	Query time / training time	Trace-ability
Pure CBR no adaptation	0.315	5.4 s / -	very good
Hybrid ANN adaptation	0.492	5.6 s / 5 s	excellent
Hybrid ANN similarity	0.453	5.4 s / 4.7 h	good
Hybrid ANN adaptation and similarity	0.516	5.6 s / 4.7 h	very good
Hybrid LWR based on CBR ($k = 10$)	0.062	5.7 s / -	below avg.
Hybrid LWR based on CBR ($k = 100$)	0.334	5.8 s / -	below avg.
Hybrid LWR based on CBR ($k = 800$)	0.532	7.0 s / -	below avg.
Pure Adaline	0.519	0.0 s / 5 s	avg.
Pure MLP	0.506	0.0 s / 10 s	none

The results of the evaluation in Table 2 confirm the hypothesis **H1**, which implies that the pure case-based approach without adaptation has lower predictive accuracy. The R^2 is considerably higher with the ANN adaptation process. The process with adaptation is only slightly slower regarding the calculation time for a query. An advantage relating to traceability is that the k most similar prices and their weights are displayed as an explanation. The reliability is increased because the relevant differences between case and query are compensated in a manner which is transparent to the user.

Additionally, the hybrid ANN similarity approach outperforms the pure CBR approach with heuristically assigned weights in terms of predictive accuracy, and thus confirms hypothesis **H2**. This advantage, however, works to the disadvantage of the other two criteria. The query time is unchanged but the ANN similarity approach has a very long training time. Moreover, the traceability of the results decreases, since in the experiments it turned out that the attributes reflecting the content of the domain received a smaller weight through the training, and thus the cases found appear less similar to the user.

Finally, the hybrid LWR approach is compared with a pure Adaline approach (10,000 iterations) and a pure MLP approach (two inner layers, each of 100 neurons, 100 iterations). The results in Table 2 confirm the hypothesis **H3** since, with 800 cases, a slightly higher predictive accuracy can be achieved with the hybrid LWR. However, the hybrid LWR has the highest query time of all the approaches presented, since the retrieval of a large number of similar cases is time-consuming and ANN training occurs for every query. Since the ANN is trained for every query, two domain names with an equal attribute value may receive a different local multiplier. In addition, the basic value (bias) is no longer the same for every domain. Hence, the hybrid LWR appraisal is very difficult to trace.

6 Conclusion and Future Work

In this paper, three hybrid approaches were introduced for the first time to domain appraisal. It has been shown that the predictive accuracy of a case-based system can be clearly increased by neural price adjustment and/or by the learning of weights. If a neural network is trained with cases in the local neighborhood of the query, then the highest predictive accuracy is achieved. For the user to have confidence in the appraisal, it is important that the method used to determine the estimated price on the basis of semantically similar cases should be traceable. Therefore, the combination of knowledge-intensive similarity measures with the adaptive character of neural networks is a key feature of the approaches presented in this paper. Depending on the application scenario (such as the mass appraisal of large domain portfolios), it is necessary to select the approach in which the criteria of traceability, predictive accuracy, and speed stand in the best relationship one to another. Genetic algorithms [15] could be used in future studies to optimize local similarity functions and local multipliers for adaptation. Parallelization with Hadoop or Amazon EC2, and improved index techniques (such as the Mac/Fac model [22] or cluster-based retrieval [23]), could in future extend the sequential retrieval. This seems particularly promising when the case base grows or when a mass appraisal of large domain portfolios needs to be performed. The applicability for car or real estate data could also be assessed.

References

1. Salvador, P., Nogueira, A.: Analysis of the Internet Domain Names Re-registration Market. In: Procedia Computer Science, 3, 325–335, Elsevier (2011)
2. Kahn, S.A., Case, F.E.: Real Estate Appraisal and Investment. The Ronald Press Company, New York (1977)
3. Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., French, N.: Real Estate Appraisal: a Review of Valuation Methods. In: Journal of Property Investment & Finance, 21(4), 383–401 (2003)
4. Phillips, G.M.: A Hedonic Regression Model for Internet Domain Name Valuation. In: Business Valuation Review, 22(2), 90–98 (2003)
5. Jindra, M.: The Market for Internet Domain Names. In: 16th European Regional Conference September 4–6, 2005 in Porto, Portugal (2005)
6. Gonzalez, A.J., Laureano-Ortiz, R.: A Case-Based Reasoning Approach to Real Estate Property Appraisal. In: Expert Systems with Applications, 4(2), 229–246, Elsevier (1992)

7. O'Roarty, B., McGreal, S., Adair, A., & Patterson, D.: Case-Based Reasoning and Retail Rent Determination. In: Journal of Property Research, 14(4), 309–328 (1997)
8. McSherry, D.: An Adaptation Heuristic for Case-Based Estimation. In: Advances in Case-Based Reasoning, 184–195, Springer (1998)
9. Dieterle, S., Bergmann, R.: Case-Based Appraisal of Internet Domains. In: Díaz-Agudo, B., Watson, I. (Eds.): Case-Based Reasoning Research and Development: 20th International Conference on Case-Based Reasoning, ICCBR 2012, 47–61, Springer (2012)
10. Samarasinghe, S.: Neural Networks for Applied Sciences and Engineering: from Fundamentals to Complex Pattern Recognition. Auerbach Publications, Boca Raton, FL (2007)
11. Rossini, P.: Application of Artificial Neural Networks to the Valuation of Residential Property. In: Third Annual Pacific-Rim Real Estate Society Conference, Palmerston North, New Zealand (1997)
12. Peterson, S., Flanagan, A.B.: Neural Network Hedonic Pricing Models in Mass Real Estate Appraisal. In: Journal of Real Estate Research, 31(2), 147–164 (2009)
13. McCluskey, W., Anand, S.: The Application of Intelligent Hybrid Techniques for the Mass Appraisal of Residential Properties. In: Journal of Property Investment & Finance, 17(3), 218–239 (1999)
14. Rossini, P.: Using Expert Systems and Artificial Intelligence for Real Estate Forecasting. In: Sixth Annual Pacific-Rim Real Estate Society Conference, Sydney, Australia, 24–27 (2000)
15. Al-Akhras, M.: An Evolutionary-Optimised Artificial Neural Network Approach for Automatic Appraisal of Jordanian Lands and Real Properties. In: NN'10/EC'10/FS'10 Proceedings of the 11th WSEAS International Conference on Neural Networks and 11th WSEAS International Conference on Evolutionary Computing and 11th WSEAS International Conference on Fuzzy Systems, 203–208 (2010)
16. Jalali, V., Leake, D.: An Ensemble Approach to Instance-Based Regression Using Stretched Neighborhoods. In: The Twenty-Sixth International FLAIRS Conference, 381–386 (2013)
17. Dieterle, S.: Neuronale und fallbasierte Bewertung von Internet-Domainnamen. Master's thesis, University of Trier (2013).
18. Warner, R.M.: Applied Statistics: From Bivariate Through Multivariate Techniques. SAGE Publications (2013)
19. Caumanns, J.: A Fast and Simple Stemming Algorithm for German Words. Technical report, Department of Mathematics and Computer Science, Free University of Berlin (1999)
20. Igel, C., Hüsken, M.: Improving the Rprop Learning Algorithm. In: Proceedings of the Second International Symposium on Neural Computation, NC 2000, 115–121, ICSC Academic Press (2000)
21. Riedmiller, M., Braun, H.: RPROP - A Fast Adaptive Learning Algorithm. In: Proceedings of ISCIS VII, University of Karlsruhe (1992)
22. Gentner, D., Forbus, K.D.: MAC/FAC: A Model of Similarity-based Retrieval. In: Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society, 504–509, Lawrence Erlbaum (1991)
23. Can, F., Altingövde, I.S., Demir, E.: Efficiency and Effectiveness of Query Processing in Cluster-Based Retrieval. In: Information Systems, 29(8), 697–717, Elsevier (2004)