

Um novo algoritmo para combinação de comitês de máquinas de Mapas Auto-Organizáveis de Kohonen de tamanhos diferentes

Leandro A. Pasa
Depto. de Engenharia Elétrica
UTFPR
Medianeira, Brasil
Email: pasa@utfpr.edu.br

José Alfredo F. Costa
Depto. de Engenharia Elétrica
UFRN
Natal, Brasil
Email: jafcosta@gmail.com

Marcial G. de Medeiros
Depto. de Engenharia Elétrica
UFRN
Natal, Brasil
Email: marcial.guerra@gmail.com

Resumo – Um comitê de máquinas é um conjunto de redes neurais trabalhando conjuntamente, cujos resultados são combinados em uma única saída, alcançando uma melhor generalização do que cada uma das redes trabalhando separadamente. Os mapas auto-organizáveis de Kohonen são redes neurais utilizadas para visualização de dados, redução de dimensionalidade e classificação de dados, entre outras aplicações. Combinar mapas auto-organizáveis de Kohonen de mesmo tamanho é uma tarefa relativamente simples. Este trabalho apresenta um método pioneiro para combinar mapas de tamanhos diferentes, com o objetivo de aumentar a acurácia de classificação, quando comparada à acurácia de um único mapa. Simulações computacionais foram realizadas com vários conjuntos de dados, demonstrando a viabilidade do método proposto.

Palavras-chave: Mapas auto-organizáveis de Kohonen, Comitês de máquinas, Índice de validação de agrupamentos.

I. INTRODUÇÃO

Um comitê de máquinas ou *ensemble* é um conjunto de redes neurais, operando independentemente, cujos resultados são combinados produzindo uma única saída, resultando em uma melhor generalização, menor variância e maior tolerância à ruídos, quando comparados com uma única rede neural [1]. É fundamental que os erros produzidos por cada um dos componentes do comitê de máquinas sejam descorrelacionados [2]. Para os mapas auto-organizáveis de Kohonen (Kohonen Self-Organizing Maps - SOM) [3], erros descorrelacionados pode ser obtidos, por exemplo, utilizando diferentes conjuntos de treinamento ou utilizando variações dos parâmetros de treinamento para cada rede neural. A combinação das saídas dos mapas SOM em uma única saída pode ser feita a partir da fusão dos neurônios dos mapas que compõem o comitê de máquinas. Esses neurônios devem representar a mesma região do espaço de entrada, ou seja: os vetores de peso dos neurônios a serem fundidos, devem ser muito próximos. Este trabalho apresenta uma metodologia para fundir mapas auto-organizáveis de Kohonen, com diferentes tamanhos, com o objetivo de aumentar a acurácia de classificação da rede.

O trabalho está organizado da seguinte maneira: as Seções II, III e IV apresentam alguns conceitos básicos e trabalhos relacionados. A Seção V apresenta o algoritmo de fusão proposto. Os resultados são exibidos e discutidos na Seção VI. A Seção VII apresenta as conclusões e algumas propostas para trabalhos futuros.

II. MAPAS AUTO-ORGANIZÁVEIS DE KOHONEN

Os mapas auto-organizáveis de Kohonen tornaram-se um modelo de rede neural muito popular. Trata-se de uma rede competitiva e de aprendizado não supervisionado que faz uma projeção não linear do espaço de entrada \mathbb{R}^p , com $p \gg 2$, em uma grade de neurônios dispostos em uma matriz bidimensional, tendo apenas duas camadas: uma de entrada de dados e outra de saída de dados.

As entradas da rede x são vetores no espaço p -dimensional. Os neurônios i da camada de saída são conectados a todas as entradas da rede, representados por um vetor de pesos sinápticos no espaço p -dimensional, $w_i = [w_{i1}, w_{i2}, \dots, w_{ip}]^T$. Estes neurônios são conectados aos neurônios adjacentes por uma relação de vizinhança que descreve a estrutura topológica do mapa.

Na fase de treinamento, em uma sequência aleatória, os padrões de entrada x são comparados com os neurônios da camada de saída. O neurônio vencedor, chamado de BMU (*Best Match Unit*), representa o vetor peso com a menor distância euclidiana em relação ao padrão de entrada x .

Chamando o neurônio vencedor de c , ele pode ser formalmente definido de acordo com a Equação 1.

$$\|x - w_c\| = \operatorname{argmin}_i \|x - w_i\| \quad (1)$$

Os pesos do BMU e dos neurônios vizinhos são ajustados de acordo com a Equação 2.

$$w_i(t+1) = w_i(t) + h_{ci}(t)[x(t) - w_i(t)] \quad (2)$$

onde t indica a interação na fase de treinamento, $x(t)$ é o padrão de entrada e $h_{ci}(t)$ é a região da vizinhança em torno do neurônio vencedor c .

Para a determinação dos tamanhos dos mapas de Kohonen, ou seja, a quantidade de neurônios dos mapas, uma abordagem descrita em [3], determina que as dimensões x e y do mapa, devem ser proporcionais aos dois maiores autovalores da matriz de covariância dos dados: λ_1 e λ_2 :

$$\frac{x}{y} \approx \sqrt{\frac{\lambda_1}{\lambda_2}} \quad (3)$$

III. ÍNDICES DE VALIDAÇÃO DE AGRUPAMENTOS MODIFICADOS

Os índices de validação de agrupamentos são utilizados para avaliar os resultados do algoritmo de agrupamento, auxiliando na tomada de decisões sobre a definição do número correto de clusters formados a partir de um conjunto de dados.

O cálculo de um índice de validação de agrupamento tem elevada complexidade computacional, o que pode ser um fator complicador em aplicações que envolvem grandes volumes de dados. Gonçalves et al. [4] propôs uma alteração nos cálculos de índices de validação, empregando a quantização vetorial produzida pelo SOM. Os vetores de pesos sinápticos (protótipos) são usados em vez dos dados originais. Desta forma, há uma diminuição da complexidade computacional para o cálculo do índice de validação.

Para contornar possíveis diferenças de valores entre o cálculo como todos os dados e com protótipos, o autor também propôs que fossem utilizados os níveis de atividade conjuntamente com os vetores de pesos sinápticos. O nível de atividade significa a quantidade de padrões de entrada que foram associados a cada protótipo pelo processo de mapeamento do SOM.

A modificação proposta no cálculo dos índices é exemplificada através do cálculo de distância entre dois agrupamentos C_i e C_j :

$$\delta_{i,j} = \frac{1}{|C_i||C_j|} \sum_{x \in C_i, y \in C_j} d(x, y) \quad (4)$$

Na Equação 4, $d(x, y)$ é uma medida de distância e $|C_i|$ e $|C_j|$ referem-se à quantidade de pontos dos clusters C_i e C_j , respectivamente. Quando a quantidade de pontos dos clusters é elevada, a complexidade computacional é elevada. A Equação 5 mostra a modificação do cálculo através da metodologia proposta:

$$\delta_{i,j}^{SOM} = \frac{1}{|C_i||C_j|} \sum_{w_i \in W_i, w_j \in W_j} h(w_i) \cdot h(w_j) \cdot d(w_i, w_j) \quad (5)$$

Onde W_i e W_j são os conjuntos de protótipos do SOM que representam os clusters C_i e C_j , respectivamente, $d(x, y)$ é a mesma medida de distância da Equação 4, $h(w_i)$ é o nível de atividade do protótipo w_i pertencente a W_i e $h(w_j)$ é o nível de atividade do protótipo w_j pertencente a W_j .

Assim o cálculo da distância apresenta um custo computacional mais baixo, uma vez que as quantidades envolvidas, W_i e W_j , são menores do que C_i e C_j . Com a inclusão dos

níveis de atividades $h(\cdot)$ dos protótipos no cálculo, obtém-se uma minimização dos erros causados pela quantização vetorial que o Mapa de Kohonen produz, já que introduz no cálculo a aproximação da densidade de pontos no espaço de entrada ora representados pelos protótipos.

IV. TRABALHOS RELACIONADOS

O interesse em aplicações de comitês de máquinas tem crescido rapidamente. É possível encontrar múltiplas variações de implementações e aplicações em diversas áreas do conhecimento.

Um trabalho utilizando fusão de mapas de Kohonen para organização e recuperação de documentos foi apresentado em [5]. Os autores compararam os resultados combinados pelo comitê de mapas de Kohonen com o SOM tradicional (um único mapa). O resultado do comitê foi superior ao resultado do mapa único.

No trabalho apresentado em [6], a combinação dos neurônios foi chamada de *Fusion-SOM* e foi baseado nos polígonos de Voronoi. O desempenho do método proposto foi superior ao SOM tradicional em termos de MSQE (*mean squared quantization error*) e preservação da topologia.

Um processo de votação ponderada, chamado WeVoS-Visom, foi descrito em [7]. O objetivo era a preservação da topologia de mapa, obtendo, consequentemente, melhor visualização dos dados. Posteriormente este algoritmo foi utilizado em um sistema híbrido para previsão de prejuízos em negócios [8].

Aplicações de comitês de mapas de Kohonen são encontradas em diversas outras áreas do conhecimento, tais como segmentação de imagens [9], robótica [10], identificação e caracterização de ataques à computadores [11], análise de dados astronômicos [12], modelos financeiros [13], entre outros.

V. MÉTODO PROPOSTO

Este trabalho apresenta um método pioneiro para combinar as saídas de mapas de Kohonen de tamanhos diferentes em uma única saída. O objetivo é melhorar a acurácia de classificação do comitê em relação à acurácia de um único mapa de Kohonen. Para este trabalho, definiu-se que o comitê será composto por sete mapas de diferentes tamanhos, derivados de um mapa base. A definição dos sete mapas é vista na descrição do algoritmo a seguir.

A. Requisitos para a combinação de mapas

No processo de fusão, os mapas são avaliados aos pares, assim como os neurônios destes mapas, com base na distância euclidiana entre eles. Mapas de tamanhos diferentes terão quantidades de neurônios diferentes. Por exemplo, se o mapa base tiver dimensões 10x10 e o mapa a ser com ele combinado tiver dimensões de 12x12, no primeiro mapa haverá 100 neurônios, enquanto que no segundo haverá 144.

Para contornar este problema, o método proposto limita a combinação da quantidade de neurônios em 100% da quantidade presente no mapa base. Assim, considerando o exemplo citado, apenas 100 neurônios, dos 144 do segundo

mapa, seriam combinados com o mapa base. Os neurônios escolhidos para a combinação são aqueles que apresentam a menor distância euclidiana em relação aos neurônios do mapa base.

B. Algoritmo de fusão

O algoritmo para a combinação dos mapas auto-organizáveis de Kohonen é assim descrito:

Entrada: Conjunto de dados.

Saída: Fusão dos mapas de Kohonen.

Passo 1: Dividir o conjunto de dados em: treinamento (80%) e teste (20%) e normalizar os dados;

Passo 2: Calcular o tamanho do Mapa Base, isto é, o mapa que será melhorado através de fusões consecutivas com outros mapas, de tamanhos diferentes, de acordo com a Equação 3;

Passo 3: A partir do Mapa Base, criar outros seis mapas, os quais diferem entre si por uma combinação de três valores, para cima e para baixo, de x e y . Considerando o Mapa Base com dimensões $N \times M$, os outros seis mapas teriam as dimensões: $(N+3) \times M$, $N \times (M+3)$, $(N-3) \times M$, $(N-3) \times (M+3)$, $(N-3) \times (M-3)$, $N \times (M-3)$, como mostrado na Fig. 1;

Passo 4: Aplicar o algoritmo SOM ao conjunto de treinamento, criando as estruturas dos sete mapas;

Passo 5: Segmentar todos os mapas, através do algoritmo k-means. Este procedimento é repetido dez vezes e o resultado escolhido é o que apresentar a menor soma dos quadrados dos erros, devido à natureza de inicialização aleatória do k-means. Este passo é supervisionado, isto é, a quantidade de clusters de cada conjunto de dados é informada ao algoritmo;

Passo 6: Calcular as performances de cada mapa. As medidas de performances utilizadas neste trabalho são o MSQE e cinco índices de validação de agrupamentos (IVA), conforme a subseção V-D;

Passo 7: Todos os mapas são ordenados de acordo com os critérios de performance, do melhor para o pior;

Passo 8: Cada mapa é sequencialmente fundido ao Mapa Base de acordo com a equação de fusão e esta fusão será ou não mantida seguindo o critério apresentado na próxima seção.

C. Combinação dos mapas

Em um comitê de redes neurais é preciso identificar e retirar do comitê as redes que não contribuem para o aumento do desempenho deste comitê [14]. Neste trabalho, o Mapa Base é o primeiro componente do comitê e os mapas restantes são candidatos a comporem o comitê. Estes mapas são organizados de acordo com a performance obtida (MSQE e os cinco índices de validação de agrupamentos). Os mapas são combinados aos pares, através da Equação 6.

$$w_c = \frac{w_i \cdot h_i + w_j \cdot h_j}{h_i + h_j} \quad (6)$$

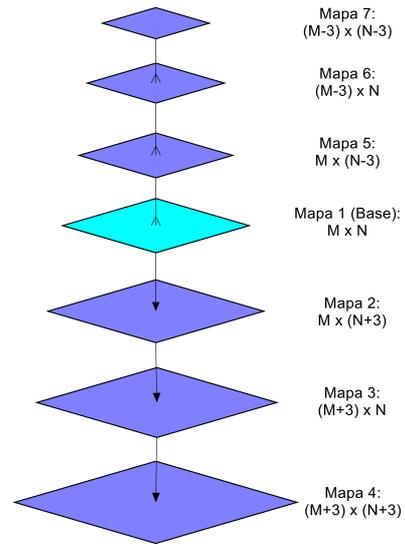


Fig. 1: Mapas criados a partir do mapa base.

onde w_c representa cada neurônio do mapa combinado; w_i e w_j são os neurônios dos mapas a serem combinados; h_i e h_j são os hits de cada neurônio de cada mapa.

O processo de combinação dos resultados de cada mapa de Kohonen inicia com o mapa base sendo combinado com o mapa que apresenta a melhor performance entre os seis outros mapas. Se a performance resultante da combinação dos mapas for aumentada em relação à performance do mapa base, a combinação destes mapas é mantida. Caso contrário, a combinação destes dois mapas é descartada e o próximo mapa, com a segunda melhor performance, é combinado com o mapa base. Cada vez que a combinação dos mapas resulta em um aumento da performance, o próximo mapa (em ordem decrescente de performance) é combinado a eles e assim sucessivamente até todos os mapas serem testados na combinação. Na próxima seção são explicadas as quatro abordagens testadas para ordenar os mapas e para estabelecer os critérios de combinação.

D. Abordagens

Nesta pesquisa, foram testadas quatro abordagens para ordenar as performances dos mapas e para estabelecer os critérios de combinação:

Abordagem 1: Mapas ordenados pelos IVA e combinados pelo critério de melhora do MSQE.

Abordagem 2: Mapas ordenados pelo MSQE e combinados pelo critério de melhora do IVA.

Abordagem 3: Mapas ordenados pelos IVA e combinados pelo critério de melhora do IVA.

Abordagem 4: Mapas ordenados pelo MSQE e combinados pelo critério de melhora do MSQE.

O MSQE (*Mean Square Quantization Error*) indica quão bem os neurônios do mapa se aproximam dos dados de entrada, ou seja, é um parâmetro para avaliar a qualidade de adaptação da rede aos dados [6]. Os índices de validação de agrupamentos empregados neste trabalho foram CD_{bw} [15],

Calinski-Harabasz [16], Dunn generalizado [17], PBM [18] e Davies-Bouldin [19], no entanto, todos eles modificados para uso com SOM, conforme proposto em [4].

VI. RESULTADOS EXPERIMENTAIS

A. Conjunto de dados

O algoritmo proposto foi testado com conjuntos de dados do Repositório UCI [20] e do FCPS (*Fundamental Clustering Problems Suite*) [21]. As características de cada conjunto de dados são mostradas na Tabela I.

Tabela I: Conjunto de dados utilizados

Dataset	Repositório	Instâncias	Atributos	Classes
BC Wisconsin	UCI	699	10	2
Column	UCI	310	6	2
Heart	UCI	303	75	2
Hepatitis	UCI	155	19	2
Ionosphere	UCI	351	34	2
Iris	UCI	150	4	3
Pima Indians	UCI	768	8	2
Seeds	UCI	210	7	3
Wine	UCI	178	13	3
Chainlink	FCPS	1000	3	2
Engytime	FCPS	4096	2	2
Lsun	FCPS	400	2	3
Tetra	FCPS	400	3	4
Two Diamonds	FCPS	800	2	2
Wingnut	FCPS	1070	2	2

B. Resultados

A Tabela II mostra os melhores resultados obtidos nas simulações. A coluna 'Índice' mostra quais índices foram melhores para cada conjunto de dados (DB refere-se ao índice Davies-Bouldin e CH refere-se ao índice Calinski and Harabasz). A coluna 'Tamanho do mapa' refere-se ao tamanho definido pela equação 3. O resultado da acurácia de classificação do comitê de mapas é comparado à acurácia de um único mapa auto-organizável de Kohonen, nas duas últimas colunas da tabela.

Com relação às abordagens definidas na seção V-D, que mostram a maneira como os mapas foram ordenados e combinados, a abordagem 1 produziu os melhores resultados para 12 dos 15 conjuntos de dados. A abordagem 2 foi melhor para os conjuntos: *Wine*, *Engytime* e *Lsun*. As abordagens 3 e 4 não produziram bons resultados de acurácia de classificação.

A Figura 2 resume os resultados de acurácia de classificação, para cada conjunto de dados, obtidos através do método proposto neste trabalho. A linha vermelha representa a acurácia para um único mapa de Kohonen e a linha verde mostra a acurácia obtida pelo comitê de mapas de Kohonen.

O teste não paramétrico de Wilcoxon (*Signed Rank Test*) [22] foi utilizado para avaliar se o resultado obtido pelo comitê de mapas de Kohonen é estatisticamente diferente das acurácias obtidas por um único mapa. A hipótese nula, H_0 , é que não há diferença entre as acurácias obtidas pelo mapa único e pelo comitê de mapas. O teste de Wilcoxon

Tabela II: Resultados experimentais

Dataset	Índice	Tamanho do mapa	Acurácia da fusão	Acurácia único mapa
BC Wisconsin	Dunn	22x6	0.9556	0.9407
Column	CH	13x10	0.7419	0.7258
Heart	CDbw	11x8	0.7119	0.6102
Hepatitis	CH	8x5	0.7333	0.5333
Ionosphere	CDbw	13x7	0.6571	0.6428
Iris	CDbw	13x5	0.9667	0.9667
Pima Indians	Dunn	17x10	0.6797	0.6797
Seeds	Dunn	14x5	0.9048	0.9048
Wine	CDbw	11x6	0.9706	0.9412
Chainlink	CH	14x11	0.7500	0.6800
Engytime	DB	20x16	0.9658	0.9646
Lsun	CDbw	11x9	0.8500	0.8000
Tetra	Dunn	11x9	0.9875	0.9875
Two Diamonds	Dunn	14x10	1.000	0.9937
Wingnut	CDbw	16x10	0.7871	0.7426

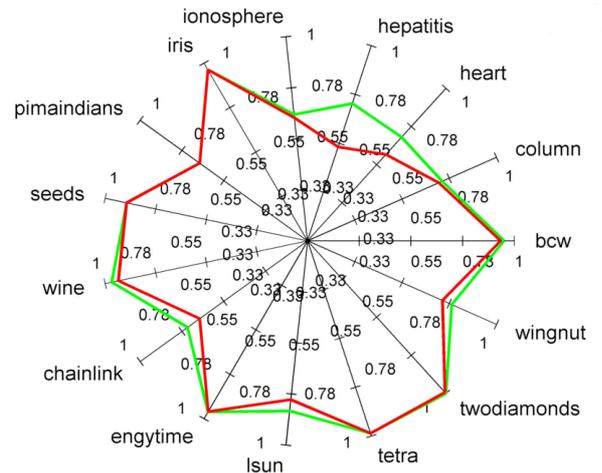


Fig. 2: Acurácia obtida.

rejeitou a hipótese nula, com valor p igual a 0.0009766, ou seja: o método proposto obteve resultados de acurácia que são diferentes (e melhores) que a acurácia de um único mapa.

VII. CONCLUSÃO

Este trabalho apresentou um método para combinar os resultados de um comitê de mapas auto-organizáveis de Kohonen, de diferentes tamanhos, com o objetivo de aumentar a acurácia de classificação em relação a uma única rede de Kohonen.

Quatro diferentes abordagens foram testadas para ordenar e combinar os mapas. A abordagem 1, em que os mapas são ordenados pelos índices de validação de agrupamentos e combinados de acordo com a redução do MSQE, mostrou-se melhor para 80% dos conjuntos de dados testados.

Embora para alguns conjuntos de dados não houve aumento na acurácia de classificação, o teste de Wilcoxon demonstrou que, no geral, o algoritmo proposto obteve resultados significativos e melhores do que um único mapa.

Trabalhos futuros incluem verificar a influência de alguns parâmetros, tais como número de mapas, variação de parâmetros de treinamento, método de segmentação dos mapas, na acurácia obtida pelo comitê de mapas auto-organizáveis de Kohonen.

REFERÊNCIAS

- [1] Dietterich, T. G.: *Ensemble methods in machine learning*. International Workshop on Multiple Classifier Systems. Springer-Verlag. 1–15, 2000.
- [2] Perrone, M. P., Cooper, L. N.: *When networks disagree: ensemble methods for hybrid neural networks*. Neural Networks for Speech and Image Processing. Chapman and Hall, 126–142, 1993.
- [3] Kohonen, T.: *Self-Organized Maps*. 2nd ed. Springer, Berlin, 1997.
- [4] Gonçalves, M.L., De Andrade Netto, M.L., Costa, J.A.F., Zullo, J.: *Data clustering using self-organizing maps segmented by mathematic morphology and simplified cluster validity indexes: an application in remotely sensed images*. IJCNN '06. International Joint Conference on Neural Networks, 4421–4428, 2006.
- [5] Georgakis, A., Li, H., Gordan, M.: *An ensemble of SOM networks for document organization and retrieval*. International Conference on Adaptive Knowledge Representation and Reasoning, 2005.
- [6] Saavedra, C., Salas, R., Moreno, S., Allende, H.: *Fusion of self-organizing maps*. 9th International Work-Conference on Artificial Neural Networks. Springer Berlin Heidelberg. 227–234, 2007.
- [7] Corchado, E., Baruque, B.: *WeVoS-ViSOM: an ensemble summarization algorithm for enhanced data visualization*. Neurocomputing. 75, 171–184, 2012.
- [8] Borrajo, M. L., Baruque, B., Corchado, E., Bajo, J., Corchado, J. M.: *Hybrid neural intelligent system to predict business failure in small-to-medium-size enterprises*. International Journal of Neural Systems 21 (04), 277–296, 2011.
- [9] Jiang, Y., Zhi-Hua, Z.: *SOM ensemble-based image segmentation*. Neural Processing Letters 20(3), 171–178, 2004.
- [10] Low, K. H., Wee, K. L., Marcelo, H. A.: *An ensemble of cooperative extended Kohonen maps for complex robot motion tasks*. Neural Computation, 17, 1411–1445, 2005.
- [11] DeLooze, L.L.: *Attack characterization and intrusion detection using an ensemble of self-organizing maps*. Information Assurance Workshop, 2006 IEEE In Information Assurance Workshop, 108–115, 2006.
- [12] Fustes, D., Dafonte, C., Arcay, B., Manteiga, M., Smith, K., Vallenari, A., Luri, X.: *SOM ensemble for unsupervised outlier analysis. Application to outlier identification in the Gaia astronomical survey*. Expert Systems with Applications 40(5), 1530–1541, 2013.
- [13] Tsai, Chih-Fong: *Combining cluster analysis with classifier ensembles to predict financial distress*. Information Fusion 16, 46–58, 2014.
- [14] Zhou, Zhi-Hua and Wu, Jianxin and Tang, Wei: *Ensembling neural networks: many could be better than all*. Artificial Intelligence. vol 137, 1-2, 239–263, 2002.
- [15] Halkidi, M. and Vazirgiannis, M.: *A density-based cluster validity approach using multi-representatives*. Pattern Recognition Letters. 29, 773–786 2008.
- [16] Milligan, G. W., Cooper, M. C. *An examination of procedures for determining the number of clusters in a data set*. Psychometrika. 50, 159–179 1985.
- [17] Bezdek, J. C., Pal, N. R. *Some new indexes of cluster validity*. IEEE Transactions Systems, Man and Cybernetic. B. 28, 301–315 1998.
- [18] Pakhira, M. K., Bandopadhyay S., Maulik U.: *Validity index for crisp and fuzzy clusters*. Pattern recognition. 37(3), 487–501 2004.
- [19] Davies, D. L., Bouldin, D. W.: *A cluster separation measure*. IEEE Transactions on Pattern Analysis and Machine Intelligence. PAMI-1, 2, 224–227 1979.
- [20] Bache, k., Lichman, M.: *Machine Learning Repository*. University of California, Irvine, School of Information and Computer Sciences, 2013.
- [21] Ultsch, A.: *Clustering with SOM: U*C*. Proceedings Workshop on Self-Organizing Maps (WSOM 2005) 75–82, 2005.
- [22] Demšar, J.: *Statistical Comparisons of Classifiers over Multiple Data Sets*. Journal of Machine Learning Research. 7, 1–30, 2006.