

Generalização do Modelo ML TSK-FS baseado na Integral de Choquet para a Classificação Multi-Etiqueta

Karina Condori¹, Julian Suarez², Giancarlo Lucca³, Graçaliz Dimuro⁴

FURG, Rio Grande, RS. Brazil

Leonardo Emmendorfer⁵

UFSM, Santa Maria, RS. Brazil

Qiongdan Lou⁶, Zhaohong Deng⁷

Jiangnan University, China

Humberto Bustince⁸

Universidad Publica de Navarra, Spain

Neste trabalho, propomos uma generalização do modelo ML-TSK FS utilizando a integral de Choquet como função de agregação, para obtenção de um novo modelo para classificação multi-etiqueta, denominado sistema fuzzy multi-etiqueta Takagi Sugeno Kang Choquet ML-TSKC FS, que é uma generalização baseada no modelo ML-TSK FS proposto por Lou [1]. O sistema fuzzy Takagi Sugeno Kang TSK é formado por k regras de inferência fuzzy que possuem uma estrutura "IF THEN", a primeira parte é chamada de antecedente e a segunda de consequente. No consequente obtemos a saída da regra normalmente uma função linear que depende dos dados de entrada e dos parâmetros de aprendizagem. A importância da regra (peso) é calculada na parte dos antecedentes, inicialmente os dados de entrada são fuzzificados para posteriormente calcular o peso. Na literatura, o operador produto é utilizado para calcular os pesos. O novo modelo é obtido substituindo o operador produto pela integral de Choquet [2]. O desempenho do ML-TSKC FS é avaliado usando a métrica Average Precision (AP) em relação a 12 conjuntos de dados multi-etiquetas obtidos do repositório MULAN. Essa avaliação é conduzida em comparação com outros 9 algoritmos de classificação multi-etiqueta. A Tabela 1 apresenta os resultados obtidos para a métrica AP.

Tabela 1: Performance dos métodos para a métrica AP

	BP-MLL	ML-kNN	CC	BR	MLSF	C2AE	JBNN	HNOML	ML-TSK FS	ML-TSKC FS
BIRDS	0.34	0.22	0.34	0.33	0.26	0.30	0.29	0.34	0.34	0.38
CAL500	0.46	0.50	0.46	0.50	0.49	0.33	0.45	0.43	0.52	0.52
FLAGS	0.82	0.80	0.80	0.81	0.82	0.74	0.80	0.81	0.82	0.83
EMOTIONS	0.80	0.71	0.78	0.80	0.76	0.57	0.76	0.80	0.82	0.82
BIBTEX	0.35	0.58	0.37	0.58	0.60	0.09	0.54	0.02	0.61	0.61
COREL	0.28	0.34	0.28	0.30	0.34	0.21	0.26	0.08	0.35	0.36
IMAGE	0.74	0.78	0.72	0.78	0.79	0.47	0.79	0.63	0.79	0.79
MIRFLICKR	0.51	0.51	0.27	0.48	0.44	0.45	0.47	0.42	0.53	0.53
REV1S1	0.49	0.61	0.52	0.57	0.60	0.21	0.53	0.05	0.61	0.62
REV1S2	0.50	0.63	0.52	0.58	0.61	0.18	0.58	0.05	0.64	0.64
SCENE	0.87	0.85	0.86	0.84	0.86	0.42	0.87	0.59	0.86	0.86
YEAST	0.76	0.61	0.75	0.72	0.76	0.57	0.74	0.71	0.76	0.77

¹kvargas@unsa.edu.pe

²jul.sejje@gmail.com

³giancarlo.lucca@furg.br

⁴gracaliz@gmail.com

⁵leonardo.emmendorfer@ufsm.br

⁶wxwangst@aliyun.com

⁷dengzhaohong@jiangnan.edu.cn

⁸bustince@unavarra.es

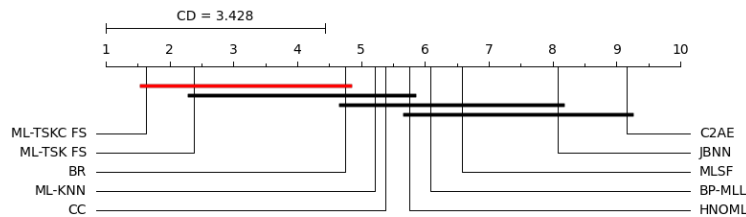


Figura 1: Comparação do ML-TSKC FS usando o teste de Bonferroni-Dunn. Fonte: O autor.

O teste de Friedman e o teste de Bonferroni-Dunn são utilizados para analisar se a diferença de desempenho entre o ML-TSKC FS e os outros métodos é estatisticamente significativa. Com base no desempenho dos métodos, conforme mostrado na Tabela 1, o teste de Friedman é realizado para a métrica AP. A hipótese nula afirma que não há diferença de desempenho entre os métodos em termos da métrica AP. Como a estatística de Friedman $F_F = 20.535$ é maior que o valor crítico 1.976, a hipótese nula é rejeitada, indicando que a diferença de desempenho observada é significativa. Em seguida, é realizado o teste post-hoc de Bonferroni-Dunn para fazer comparações entre o ML-TSKC FS e os outros métodos. A diferença entre o ranking médio do ML-TSKC FS e de outro método é expressa em termos da diferença crítica (CD). Com um nível de confiança $\alpha = 0.05$ e $q_\alpha = 2.773$, temos $CD = 3.428$. A Figura 1 apresenta os rankings médios dos 10 métodos, dispostos da esquerda (mais superior) para a direita em ordem ascendente. Quanto menor o valor do ranking, mais superior é o método. Se a diferença do ranking médio entre o ML-TSKC FS e um método for inferior a um CD, a diferença do ranking é considerada não significativa. Caso contrário, a diferença é significativa. Na Figura 1, o ML-TSKC FS está conectado a um método por meio de uma linha vermelha quando não há diferença significativa no ranking médio entre eles. A partir da Figura 1, podemos obter as seguintes conclusões: O ML-TSKC FS é significativamente superior ao ML-kNN, CC, C2AE, JBNN, MLSF, BP-MLL e HNOML em termos de AP. Embora o desempenho do ML-TSKC FS não seja significativamente diferente do ML-TSK FS e do BR, o ML-TSKC FS supera esses dois métodos, como evidenciado pelos resultados na Tabela 1.

Agradecimentos

Esta pesquisa é financiada pela FAPERGS/Brasil (Proc. 19/2551-0001660-3), CNPq/Brasil (301618/2019-4, 305805/2021-5), Agencia Estatal de Investigación (Governo da Espanha) no âmbito do projeto PID2019 -108392GB-I00 (AEI/10.13039/501100011033), CNPq/FAPERGS (Edital 07/2022 - Programa de Apoio à Fixação de Jovens Doutores no Brasil), CNPq/FAPERGS (Edital 23/2551-0000126-8).

Referências

- [1] Qiongdan L., Zhaohong D., Zhiyong X., Kup-Sze C. e Shitong W. “Multilabel Takagi-Sugeno-Kang Fuzzy System”. Em: **IEEE Transactions on Fuzzy Systems** 30.9 (set. de 2022), pp. 3410–3425.
- [2] G. Lucca, J. A. Sanz, G. P. Dimuro, E. N. Borges, H. Santos e H. Bustince. “Analyzing the performance of different fuzzy measures with generalizations of the Choquet integral in classification problems”. Em: **2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)**. IEEE. 2019, pp. 1–6.