



Wrocław
University
of Technology

**Social Network Group
@ Wrocław University of Technology**



Social Network Group
@ Wrocław University of Technology

Zaproszenie na spotkanie

Data i godzina: 2012-05-15, godz. 10:00

Miejsce: Politechnika Wrocławska, bud. A-1, I p., s. 202 (wejście przez s. 203)

Agenda:

10:00-11:00 – Tomasz Filipowski (Instytut Informatyki Politechniki Wrocławskiej):

“Correcting evaluation bias of relational classifiers with network cross validation”

11:15-12:15 – Andrzej Jarynowski (Instytut Fizyki, Uniwersytet Jagielloński):

“Sieci społeczne w modelowaniu epidemiologicznym”

Po prezentacjach planujemy wspólny obiad.

Nasza strona:

<http://www.ii.pwr.wroc.pl/~sna/>

Pozostałe informacje:

Zapraszamy na:

- **SNAA 2012** – The Second Workshop on Social Network Analysis in Applications - <http://www.zsi.pwr.wroc.pl/SNAA2012/>
- **CSNA 2012** – International Workshop on Complex Social Network Analysis - <http://www.zsi.pwr.wroc.pl/CSNA2012/>

Następne strony zawierają krótkie wprowadzenia do jednego z tematów prezentowanych na najbliższym spotkaniu (tzw. *one-pagers*).

Tomasz Filipowski (Politechnika Wrocławska):

Correcting evaluation bias of relational classifiers with network cross validation

Podczas prezentacji zostaną omówione koncepcje przedstawione w artykule "*Correcting evaluation bias of relational classifiers with network cross validation*" (Jennifer Neville, Brian Gallagher, Tina Eliassi-Rad and Tao Wang, Knowledge and Information Systems, vol. 30, pp. 31-55, Springer, 2012).

URL do artykułu

<http://dx.doi.org/10.1007/s10115-010-0373-1>

Abstrakt

Recently, a number of modeling techniques have been developed for data mining and machine learning in relational and network domains where the instances are not independent and identically distributed (i.i.d.). These methods specifically exploit the statistical dependencies among instances in order to improve classification accuracy. However, there has been little focus on how these same dependencies affect our ability to draw accurate conclusions about the performance of the models. More specifically, the complex link structure and attribute dependencies in relational data violate the assumptions of many conventional statistical tests and make it difficult to use these tests to assess the models in an unbiased manner. In this work, we examine the task of within-network classification and the question of whether two algorithms will learn models that will result in significantly different levels of performance. We show that the commonly used form of evaluation (paired t-test on overlapping network samples) can result in an unacceptable level of Type I error. Furthermore, we show that Type I error increases as (1) the correlation among instances increases and (2) the size of the evaluation set increases (i.e., the proportion of labeled nodes in the network decreases). We propose a method for network cross-validation that combined with paired t-tests produces more acceptable levels of Type I error while still providing reasonable levels of statistical power (i.e., 1–Type II error).