CS 512

Assignment 2

Bing Liang NetID: bingl3

Result

	F1 (micro)	F1 (macro)
text-based features	0.28940823231531065	0.192858514599255
HIN-based features	0.9827018306482332	0.7579009357983942

Analysis

According to the F1 score (both micro and macro), we can find that the performance of using HIN-based features is far more better using text-based features. I believe the reasons behind it is: the text-based features is kind like a one-hot encoding, which may very sparse for training the classifier model. Turns out, the HIN-based model could utilize the meta-path information which make it have a better way to represent the features of a paper. And we could learn from this result, the cited venue information is more important than paper's title. It is make sense because intuitively, if a paper cites a paper form a venue, it means this paper has very high probability has similar focus area to the cited paper, as a result, it would be a good suggestion to publish on same venue.

Further Improvement

In this task, we only utilize one type of meta-path, but actually we could utilize more types of meta-path. For example, we could find some better ways to represents the features, like

Metapath2vec to learn the embedding of a meta-path. Furthermore, we could consider each relation in metapath entities as a neural net, and use deep learning way to learn the embedding, which is our research project doing.

Precision and Recall Per Venue

Text-based features (could also be found in output/result_simple_clf.txt):

Venue Precision Recall

aaai 0.08236151603498543 0.21401515151515152

aamas 0.35294117647058826 0.25787965616045844

acc 0.0 0.0

acm_multimedia 0.2593856655290102 0.20052770448548812

acm_trans._graph. 0.0 0.0

amcis 0.32673267326732675 0.24812030075187969

amia 0.5094339622641509 0.4462809917355372

asp-dac 0.3448275862068966 0.03134796238244514

bioinformatics 0.0 0.0

cdc 0.46866485013623976 0.6120996441281139

chi 0.29831932773109243 0.16985645933014354

chi_extended_abstracts 0.30851063829787234 0.4084507042253521

cikm 0.211864406779661 0.06631299734748011

cogsci 0.47276688453159044 0.5

coling 0.385 0.2098092643051771

commun._acm 0.08433734939759036 0.10294117647058823

compsac 0.33766233766233766 0.08783783783783784

comput._graph._forum 0.0 0.0

comput._j. 0.0 0.0

computer_communications 0.0 0.0

computer_networks 0.0 0.0

corr 0.0 0.0

cvpr 0.23679060665362034 0.20404721753794267

dac 0.39325842696629215 0.15695067264573992

date 0.19902912621359223 0.1673469387755102

ecai 0.12121212121212122 0.0213903743315508

ecis 0.33031674208144796 0.18961038961038962

embc 0.45454545454545453 0.24853801169590642

encyclopedia_of_database_systems 0.28150765606595995 0.7940199335548173

etfa 0.44761904761904764 0.1843137254901961

eurospeech 0.25 0.0220125786163522

eusipco 0.1151603498542274 0.09272300469483569

expert_syst._appl. 0.0 0.0

focs 0.20238095238095238 0.06563706563706563

fskd 0.23076923076923078 0.031141868512110725

fundam._inform. 0.0 0.0

fusion 0.48792270531400966 0.3531468531468531

fuzz-ieee 0.5124610591900312 0.6714285714285714

gecco 0.5329341317365269 0.23116883116883116

globecom 0.14878892733564014 0.16250944822373395

hicss 0.37667560321715815 0.28644240570846075

icalt 0.47230320699708456 0.47928994082840237

icarcv 0.0 0.0

icassp 0.17696629213483145 0.26534296028880866

icc 0.16281512605042017 0.2246376811594203

iccad 0.23846153846153847 0.10264900662251655

iccs 0.3523809523809524 0.17535545023696683

iccv 0.08823529411764706 0.011029411764705883

icdar 0.5521978021978022 0.7077464788732394

icde 0.20207253886010362 0.11370262390670553

icecs 0.21875 0.023890784982935155

icip 0.2707509881422925 0.2893347412882788

icis 0.23893805309734514 0.06633906633906633

icmc 0.5008156606851549 0.7057471264367816

icme 0.2261904761904762 0.03518518518518519

icml 0.17391304347826086 0.045454545454545456

icnc 0.36363636363636365 0.05714285714285714

icpr 0.13636363636363635 0.019736842105263157

icra 0.4139365574622985 0.44394868934746234

icse 0.3423728813559322 0.3117283950617284

icslp 0.38095238095238093 0.028268551236749116

ieee_computer 0.0 0.0

ieee_congress_on_evolutionary_computation 0.4036697247706422 0.4444444444444444

ieee_journal_on_selected_areas_in_communications 0.0 0.0

ieee_software 0.0 0.0

ieee_trans._computers 0.0 0.0

ieee_trans._information_theory 0.0 0.0

ieee_trans._knowl._data_eng. 0.0 0.0

ieee_trans._parallel_distrib._syst. 0.0 0.0

ieee_trans._pattern_anal._mach._intell. 0.0 0.0

ieee_trans._software_eng. 0.0 0.0

igarss 0.72 0.8694339622641509

ijcai 0.1061662198391421 0.2632978723404255

ijcnn 0.34563758389261745 0.18968692449355432

inf._process._lett. 0.0 0.0

inf._sci. 0.0 0.0

infocom 0.2697768762677485 0.1927536231884058

int._cmg_conference 0.5265486725663717 0.664804469273743

interspeech 0.4532299741602067 0.700479233226837

ipdps 0.3552123552123552 0.20489977728285078

iros 0.32695810564663025 0.2566118656182988

isbi 0.5186813186813187 0.565947242206235

iscas 0.3630069238377844 0.4128233970753656

iscc 0.056910569105691054 0.021671826625386997

isit 0.384012539184953 0.532608695652174

itc 0.5448275862068965 0.6991150442477876

j._acm 0.0 0.0

j._parallel_distrib._comput. 0.0 0.0

j._symb._log. 0.0 0.0

journal_of_systems_and_software 0.0 0.0

kdd 0.24778761061946902 0.11715481171548117

lcn 0.30952380952380953 0.045454545454545456

lrec 0.521484375 0.579175704989154

multimedia_tools_appl. 0.0 0.0

neuroimage 0.0 0.0

nips 0.2226148409893993 0.21428571428571427

pacis 0.2823529411764706 0.0821917808219178

pattern_recognition 0.0 0.0

pdpta 0.3484848484848485 0.07641196013289037

pimrc 0.20350404312668463 0.16815144766146994

robio 0.2543352601156069 0.10891089108910891

sac 0.12224108658743633 0.12478336221837089

siam_j._comput. 0.0 0.0

sigcse 0.6302325581395349 0.6913265306122449

sigir 0.33440514469453375 0.2962962962962963

sigmod_conference 0.3146853146853147 0.1505016722408027

smc 0.1111111111111111 0.07577497129735936

soda 0.28378378378378377 0.2876712328767123

softw.,_pract._exper. 0.0 0.0

stoc 0.2885375494071146 0.2347266881028939

theor._comput._sci. 0.07142857142857142 0.06666666666666667

vlsi_design 0.19230769230769232 0.03690036900369004

vtc_fall 0.12598425196850394 0.061420345489443376

vtc_spring 0.16393442622950818 0.033112582781456956

wcnc 0.1487603305785124 0.046094750320102434

winter_simulation_conference 0.5655430711610487 0.6918671248568156

HIN-based features (could also be found in output/result_hin_clf.txt):

Venue Precision Recall

aamas 1.0 1.0

acc 1.0 1.0

acm_multimedia 1.0 1.0

acm_trans._graph. 0.0 0.0

amcis 1.0 1.0

amia 1.0 1.0

asp-dac 1.0 1.0

bioinformatics 1.0 1.0

cdc 1.0 1.0

chi 1.0 1.0

chi_extended_abstracts 1.0 1.0

cikm 0.9189944134078212 0.8726790450928382

cogsci 1.0 1.0

coling 1.0 1.0

commun._acm 0.4666666666666667 0.10294117647058823

compsac 1.0 1.0

comput._graph._forum 0.0 0.0

comput._j. 0.2 0.05263157894736842

computer_communications 0.0 0.0

computer_networks 0.11111111111111110.5

corr 0.0 0.0

cvpr 0.97627737226277370.9021922428330523

dac 0.99775784753363230.9977578475336323

date 1.0 1.0

ecai 0.95811518324607330.9786096256684492

ecis 1.0 1.0

embc 1.0 1.0

encyclopedia_of_database_systems 0.84550561797752811.0

etfa 1.0 1.0

eurospeech 1.0 1.0

eusipco 1.0 1.0

expert_syst._appl. 0.0 0.0

focs 1.0 1.0

fskd 1.0 1.0

fundam._inform. 0.0 0.0

fusion 0.99305555555555561.0

fuzz-ieee 1.0 1.0

gecco 1.0 1.0

globecom 1.0 1.0

hicss 1.0 1.0

icalt 1.0 1.0

icarcv 1.0 0.9962825278810409

icassp 1.0 1.0

icc 1.0 1.0

iccad 1.0 1.0

iccs 1.0 0.995260663507109

iccv 0.88842975206611570.7904411764705882

icdar 1.0 1.0

icde 0.790625 0.7376093294460642

icecs 1.0 1.0

icip 1.0 1.0

icis 1.0 1.0

icmc 1.0 1.0

icme 1.0 1.0

icml 0.97637795275590560.9393939393939394

icnc 1.0 1.0

icpr 1.0 1.0

icra 1.0 1.0

icse 0.99692307692307691.0

icslp 1.0 1.0

ieee_computer0.25 0.05555555555555555

ieee_congress_on_evolutionary_computation 1.0 1.0

ieee_journal_on_selected_areas_in_communications 0.0 0.0

ieee_software 0.0 0.0

ieee_trans._computers 0.0 0.0

ieee_trans._information_theory 0.0 0.0

ieee_trans._knowl._data_eng. 0.0 0.0

ieee_trans._parallel_distrib._syst. 0.0 0.0

ieee_trans._pattern_anal._mach._intell. 0.0 0.0

ieee_trans._software_eng. 0.29411764705882354 0.10869565217391304

igarss 1.0 0.999245283018868

ijcai 1.0 1.0

ijcnn 1.0 1.0

inf._process._lett. 0.0 0.0

inf._sci. 0.0 0.0

infocom 1.0 1.0

int._cmg_conference 1.0 0.9972067039106145

interspeech 1.0 1.0

ipdps 1.0 1.0

iros 1.0 1.0

isbi 1.0 1.0

iscas 0.7943327239488117 0.9775028121484814

iscc 1.0 1.0

isit 1.0 1.0

itc 0.9480812641083521 0.9292035398230089

j._acm 0.12121212121212122 0.125

j._parallel_distrib._comput. 0.0 0.0

j._symb._log. 0.0 0.0

journal_of_systems_and_software 0.0 0.0

kdd 0.99581589958159 0.99581589958159

lcn 1.0 1.0

lrec 1.0 1.0

multimedia_tools_appl. 0.0 0.0

neuroimage 0.0 0.0

nips 1.0 1.0

pacis 1.0 1.0

pattern_recognition 0.6 1.0

pdpta 1.0 1.0

pimrc 1.0 1.0

robio 1.0 1.0

sac 1.0 1.0

siam_j._comput. 1.0 0.07142857142857142

sigcse 1.0 1.0

sigir 0.9077809798270894 0.8974358974358975

sigmod_conference 0.8006872852233677 0.7792642140468228

smc 0.9885321100917431.0

soda 1.0 0.9965753424657534

softw._pract._exper. 0.0 0.0

stoc 1.0 0.9935691318327974

theor._comput._sci. 0.0 0.0

vlsi_design 1.0 0.996309963099631

vtc_fall 1.0 1.0

vtc_spring 1.0 1.0

wcnc 1.0 1.0

winter_simulation_conference 1.0 1.0