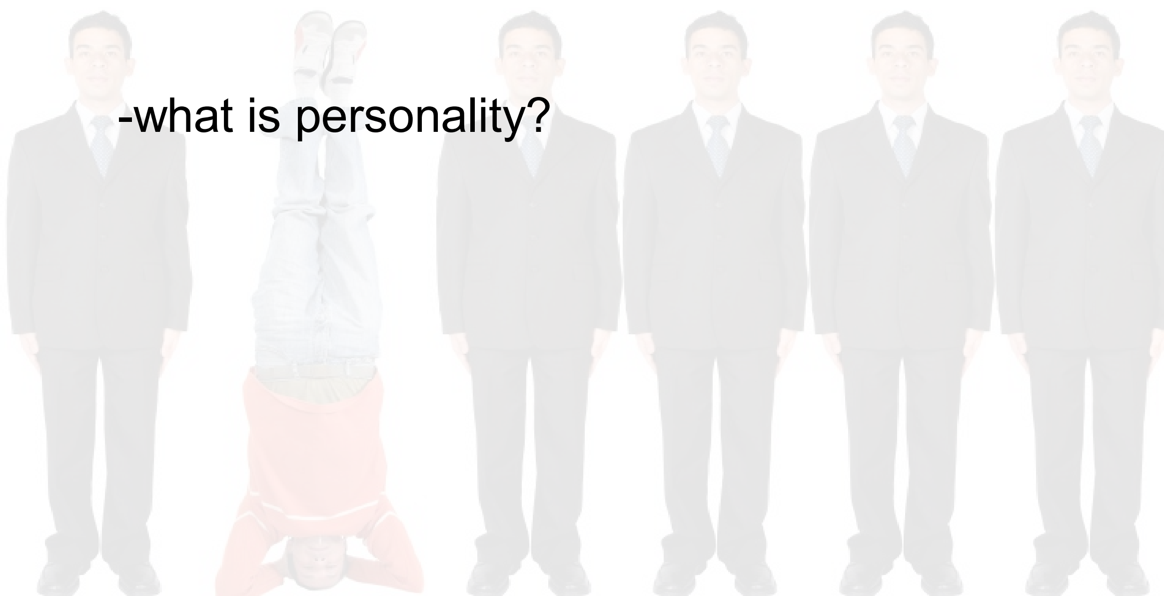
Fabio Celli
fabio.celli@unitn.it

Unsupervised Personality Recognition from Text: Possible Applications

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

Unsupervised Personality Recognition from Text:
Possible Applications



-what is personality?

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

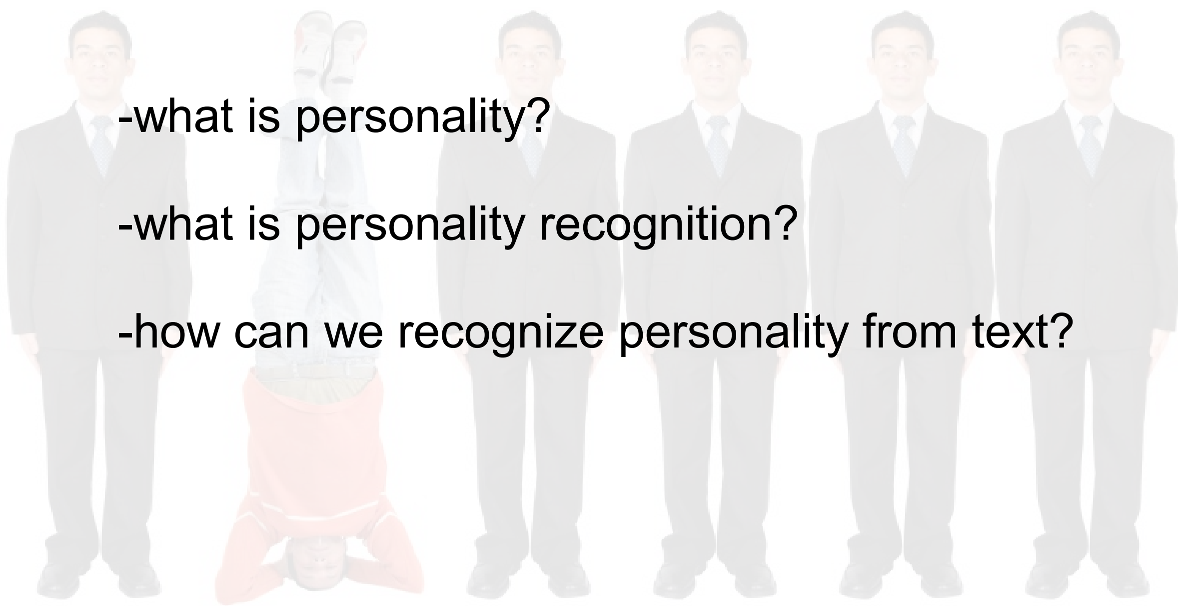
Unsupervised Personality Recognition from Text: Possible Applications

-what is personality?

-what is personality recognition?

Fabio Celli
fabio.celli@unitn.it

Unsupervised Personality Recognition from Text: Possible Applications



-what is personality?

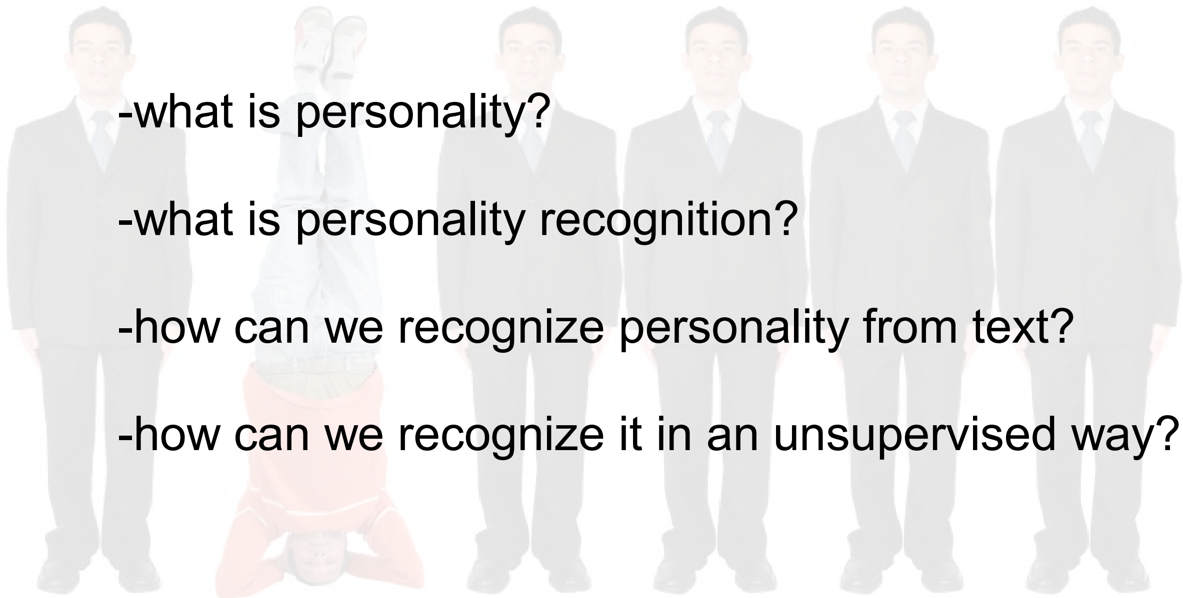
-what is personality recognition?

-how can we recognize personality from text?

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

Unsupervised Personality Recognition from Text: Possible Applications



-what is personality?

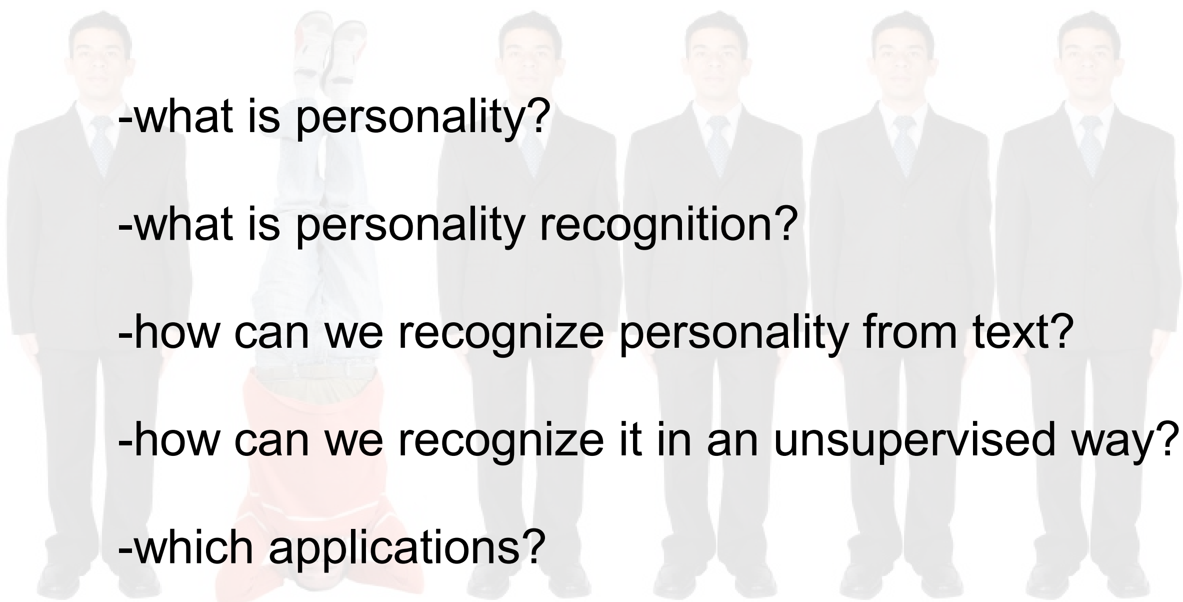
-what is personality recognition?

-how can we recognize personality from text?

-how can we recognize it in an unsupervised way?

Fabio Celli
fabio.celli@unitn.it

Unsupervised Personality Recognition from Text: Possible Applications



-what is personality?

-what is personality recognition?

-how can we recognize personality from text?

-how can we recognize it in an unsupervised way?

-which applications?

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

“ Personality describes persistent human behavioral responses to broad classes of environmental stimuli.
[Adelstein *et al* 2011] ”

Fabio Celli
fabio.celli@unitn.it

The Myers-Briggs ™ Type Indicator

(The Keirsey Temperament Sorter)

E Extroverted (Expressive)	**S** Sensing (Observant)	**T** Thinking (Tough-Minded)	**J** Judging (Scheduling)
I Introverted (Reserved)	**N** Intuitive (Introspective)	**F** Feeling (Friendly)	**P** Perceiving (Probing)

The Big 5 factor theory

extrovert	**shy**
stable	**neurotic**
friendly	**uncooperative**
organized	**careless**
insightful	**unimaginative**

Fabio Celli
fabio.celli@unitn.it

The Myers-Briggs ™ Type Indicator

(The Keirsey Temperament Sorter)

E Extroverted (Expressive)	**S** Sensing (Observant)	**T** Thinking (Tough-Minded)	**J** Judging (Scheduling)
I Introverted (Reserved)	**N** Intuitive (Introspective)	**F** Feeling (Friendly)	**P** Perceiving (Probing)

The Big 5 factor theory

extrovert	**shy**
stable	**neurotic**
friendly	**uncooperative**
organized	**careless**
insightful	**unimaginative**

-self assessments
-observed assessments (+agreement)

Fabio Celli
fabio.celli@unitn.it

The Myers-Briggs ™ Type Indicator

(The Keirsey Temperament Sorter)

E Extroverted (Expressive)	**S** Sensing (Observant)	**T** Thinking (Tough-Minded)	**J** Judging (Scheduling)
I Introverted (Reserved)	**N** Intuitive (Introspective)	**F** Feeling (Friendly)	**P** Perceiving (Probing)

The Big 5 factor theory

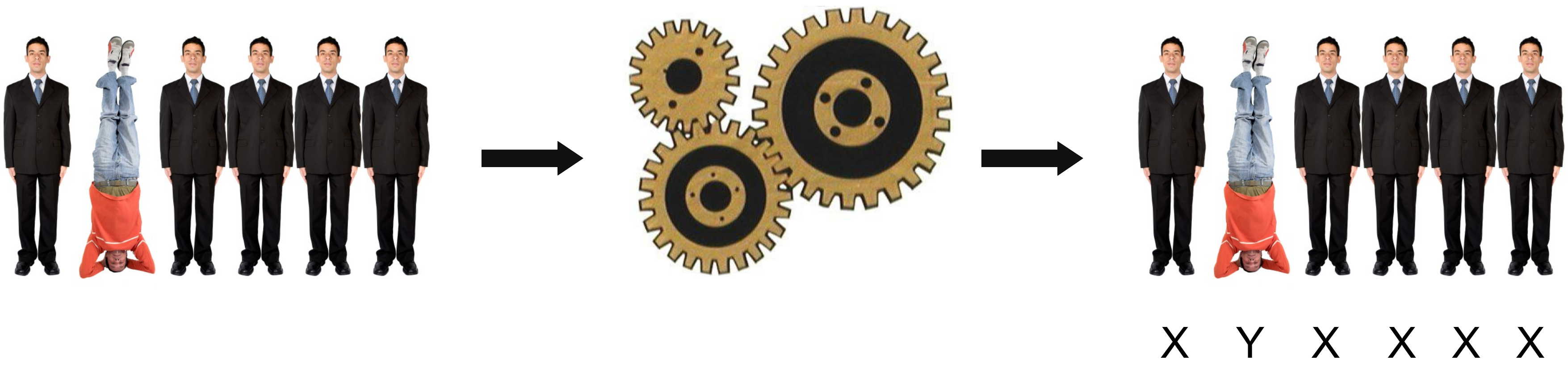
extrovert	**shy**
stable	**neurotic**
friendly	**uncooperative**
organized	**careless**
insightful	**unimaginative**

-self assessments
-observed assessments (+agreement)

-100 item test
-50 item test
-44 item test
-10 item test

Ground truth

Fabio Celli
fabio.celli@unitn.it



personality recognition

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

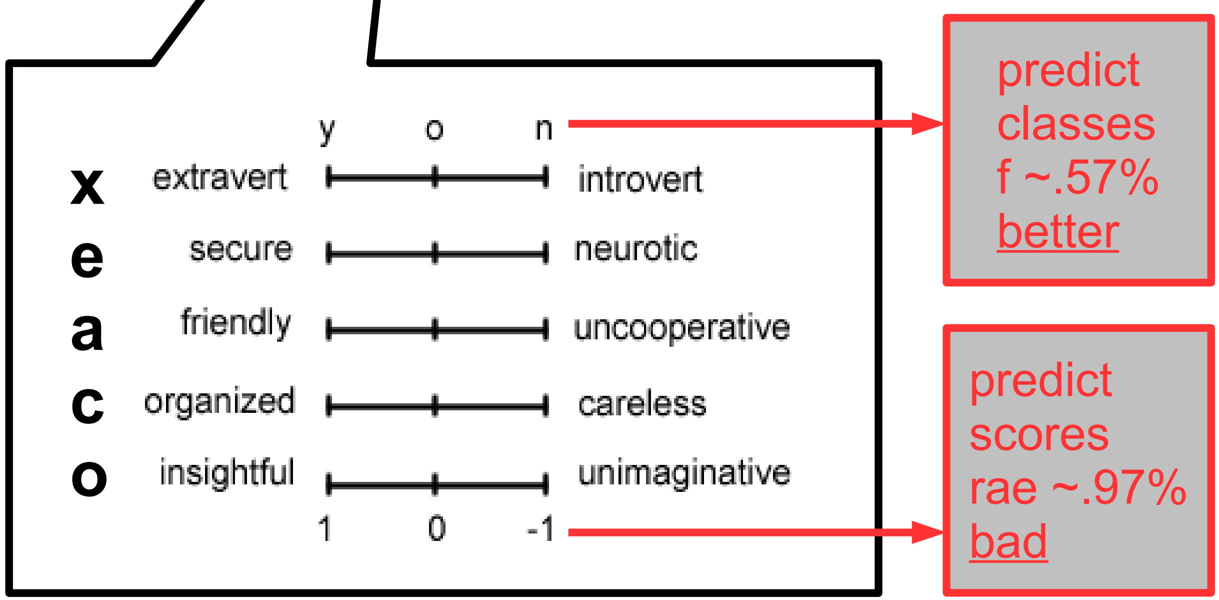
Personality Recognition

is the automatic classification of the personality of authors from behvioral features (text, facial expressions, profile pictures, works, and so on). gold standard labels can be obtained by means of the big5 personality tests.
[Norman 1963; Costa & MacRae 1985; Digman 1990]

Fabio Celli
fabio.celli@unitn.it

Personality Recognition

is the automatic classification of the personality of authors from behvioral features (text, facial expressions, profile pictures, works, and so on). gold standard labels can be obtained by means of the big5 personality tests.
[Norman 1963; Costa & MacRae 1985; Digman 1990]



[Mairesse et Al. 2007]

Fabio Celli
fabio.celli@unitn.it

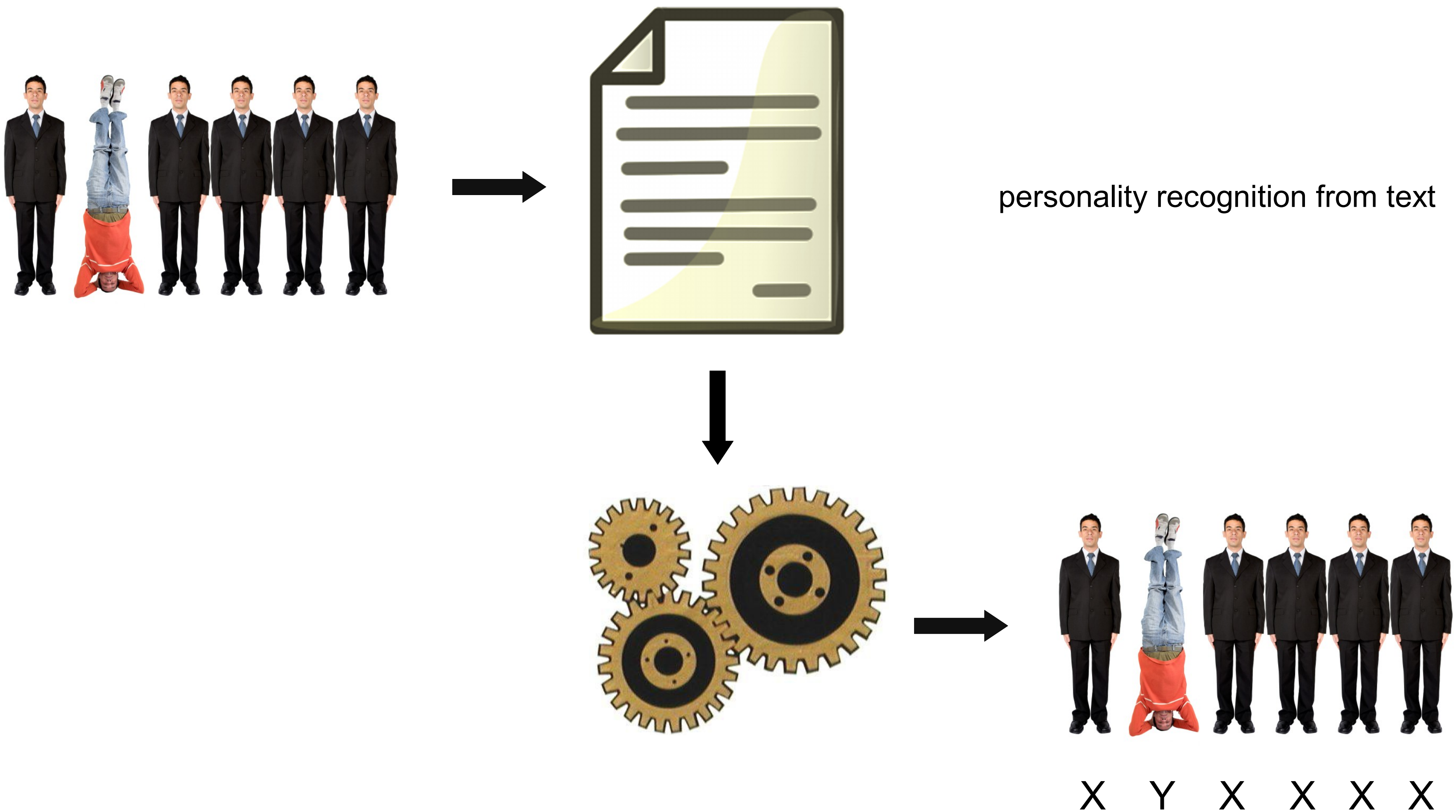
Personality Recognition

is the automatic classification of the personality of authors from behvioral features (text, facial expressions, profile pictures, works, and so on). gold standard labels can be obtained by means of the big5 personality tests.
[Norman 1963; Costa & MacRae 1985; Digman 1990]

5 classifiers (one per trait)

predict binary classes or scores

Fabio Celli
fabio.celli@unitn.it



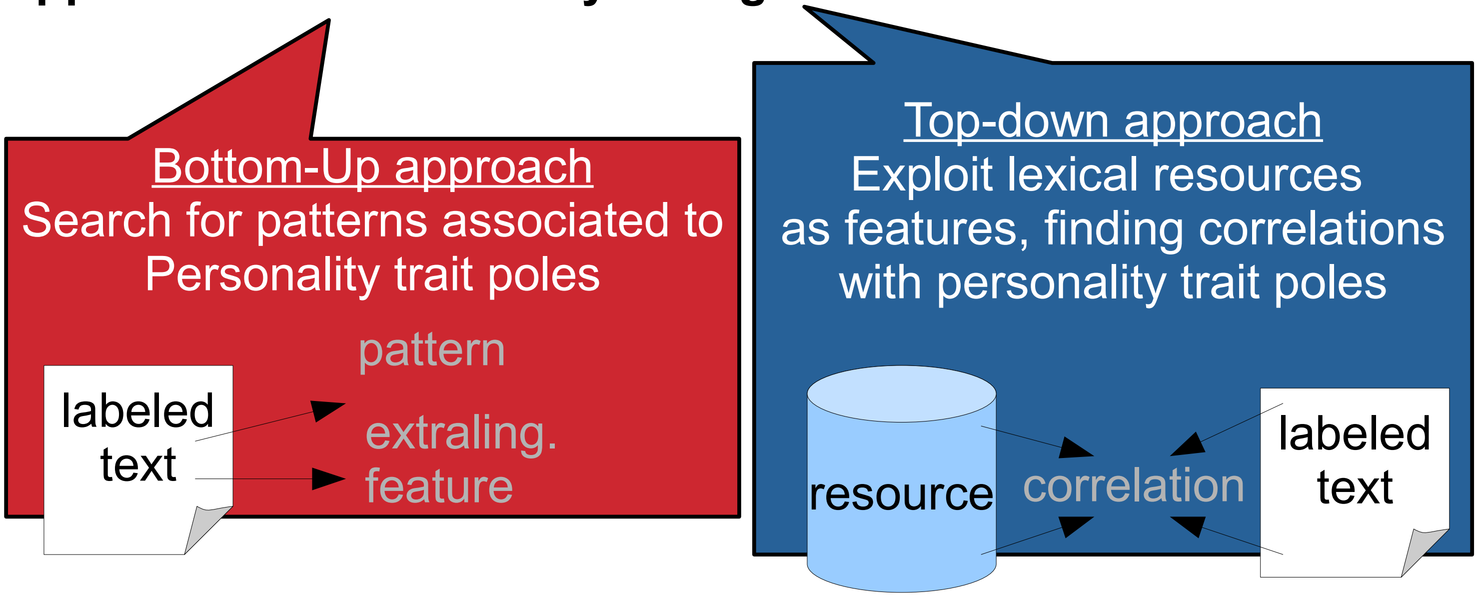
Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

Approaches to Personality Recognition from text

[Oberlander & Nowson 2006]

[Iacobelli *et al* 2011]



[Mairesse *et al* 2007]

[Scwartz *et al* 2013]

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

Approaches to Personality Recognition from text

[Oberlander & Nowson 2006]

[Iacobelli *et al* 2011]

Bottom-Up approach
Search for patterns associated to Personality trait poles

labeled text
pattern
extraling. feature

Top-down approach
Exploit lexical resources as features, finding correlations with personality trait poles

resource
correlation
labeled text

[Mairesse *et al* 2007]

[Scwartz *et al* 2013]

[Markovikj *et al* 2013]

Mixed approach
Use many resources (sentiment, Psycholinguistic, semantic) + word patterns + feature selection

Sep 16, 2014

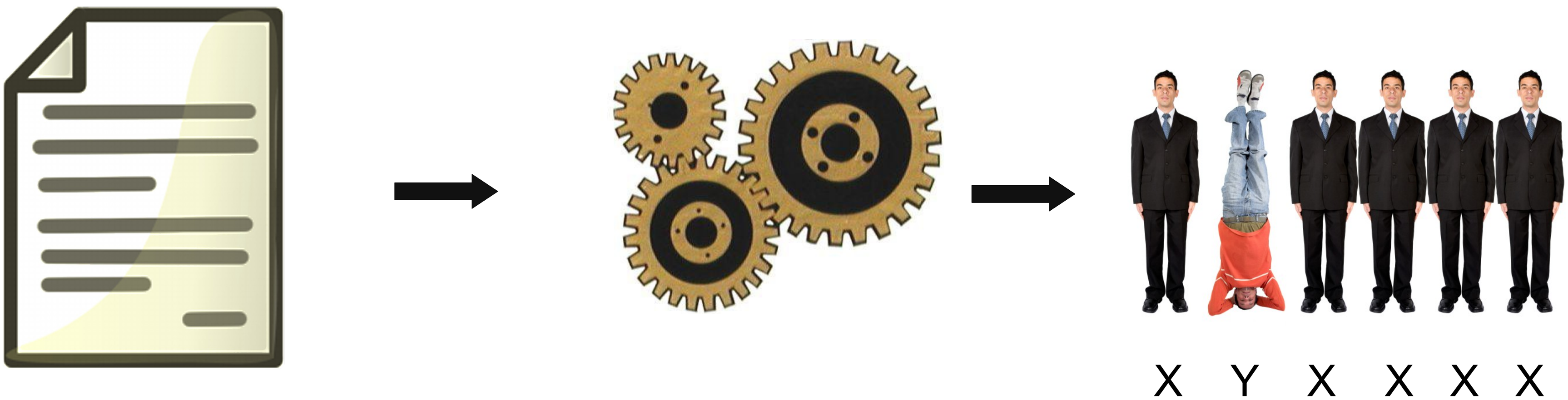
Approaches to Personality Recognition from text

5 classifiers (one per trait)

predict binary classes or scores

Large feature space, reuced with feature selction

Fabio Celli
fabio.celli@unitn.it

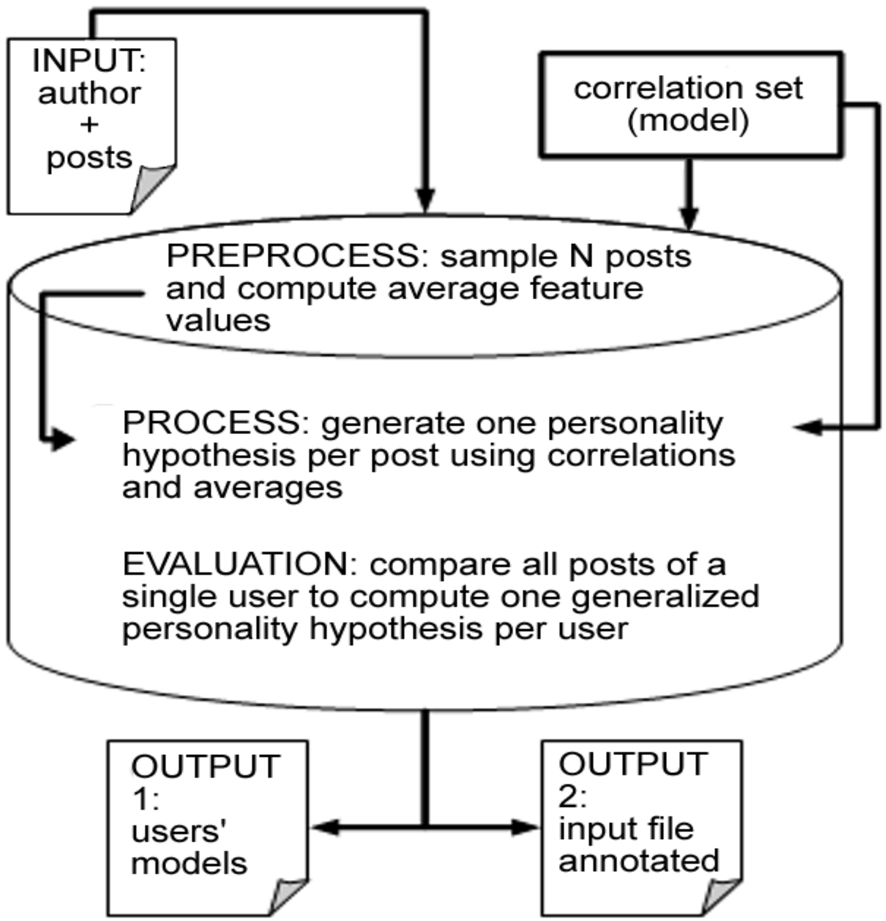


Unsupervised personality recognition from text

Sep 16, 2014

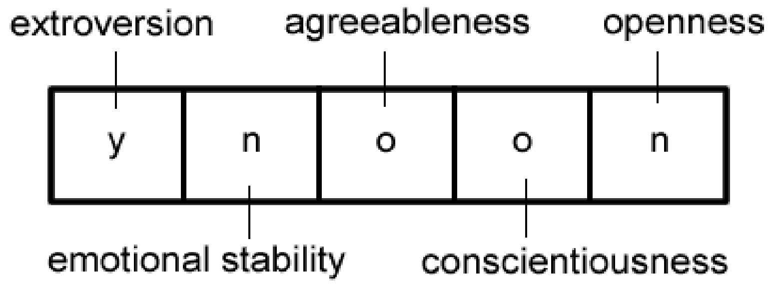
Fabio Celli
fabio.celli@unitn.it

Unsupervised personality recognition from text



We need:

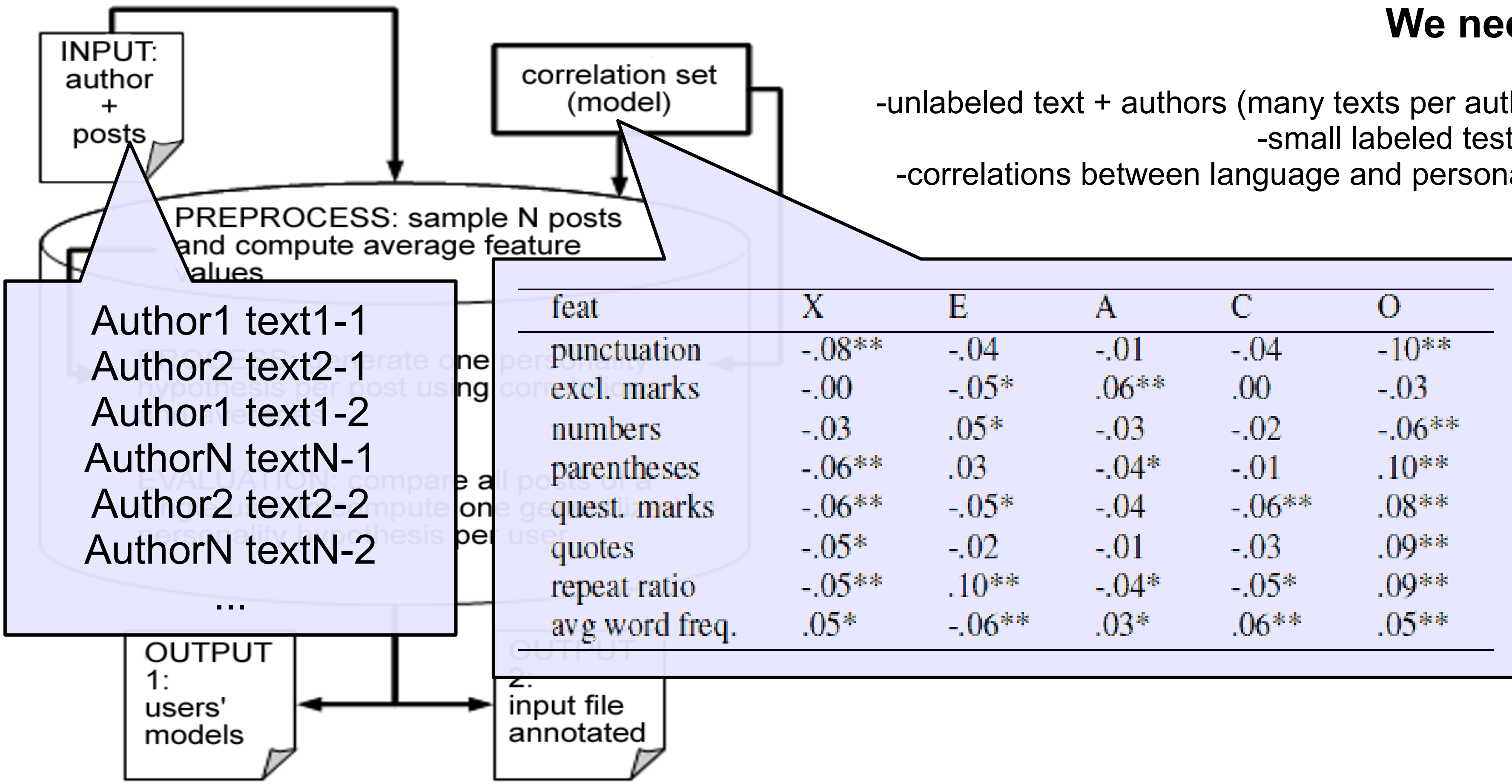
-unlabeled text + authors (many texts per author)
-small labeled test set
-correlations between language and personality



In literature:
3 classes: high, (y) mid, (o) low (n)
2 classes: high (y) low (n)

Fabio Celli
fabio.celli@unitn.it

Unsupervised personality recognition from text

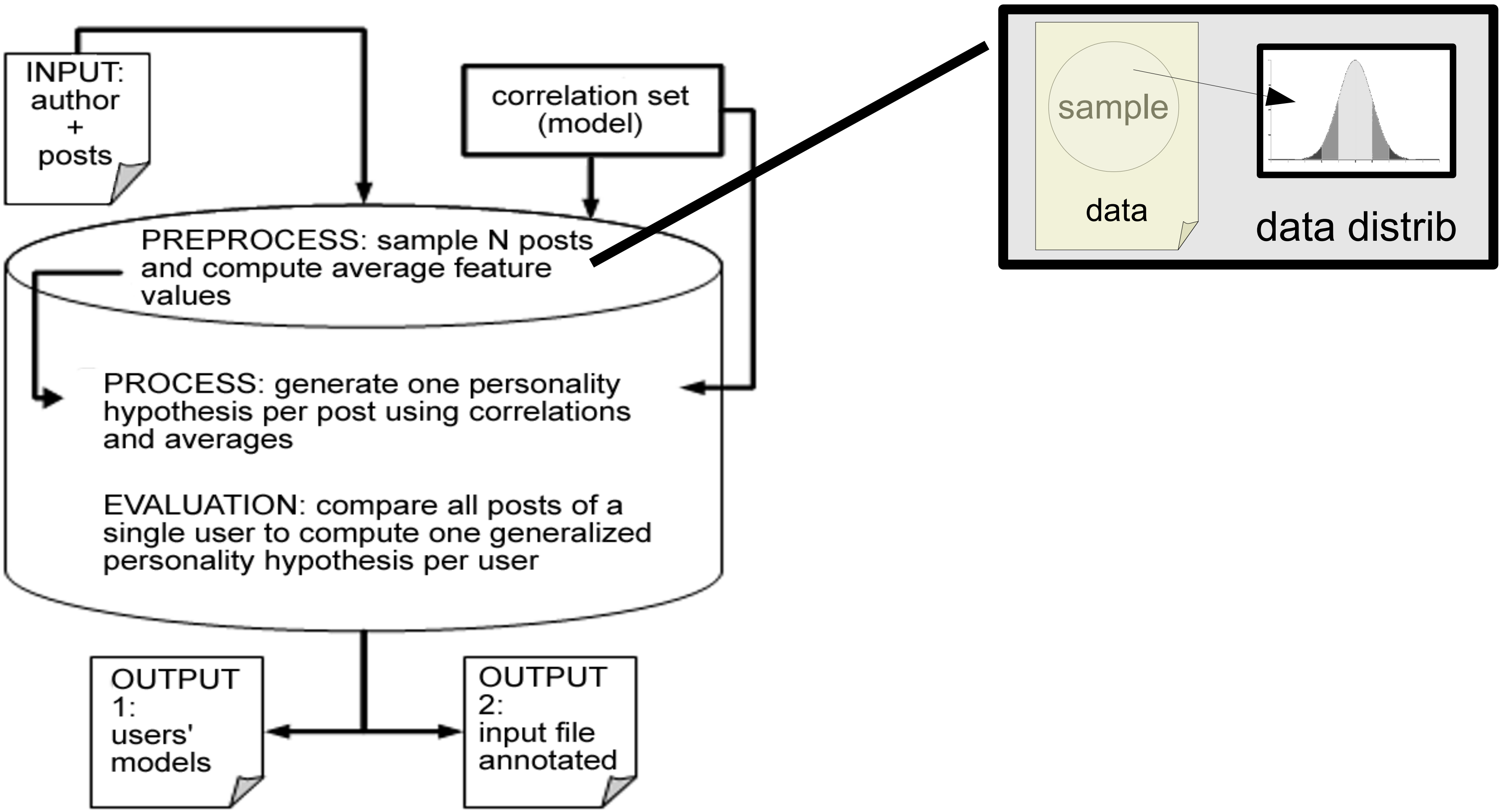


feat	X	E	A	C	O
punctuation	-.08**	-.04	-.01	-.04	-10**
excl. marks	-.00	-.05*	.06**	.00	-.03
numbers	-.03	.05*	-.03	-.02	-.06**
parentheses	-.06**	.03	-.04*	-.01	.10**
quest. marks	-.06**	-.05*	-.04	-.06**	.08**
quotes	-.05*	-.02	-.01	-.03	.09**
repeat ratio	-.05**	.10**	-.04*	-.05*	.09**
avg word freq.	.05*	-.06**	.03*	.06**	.05**

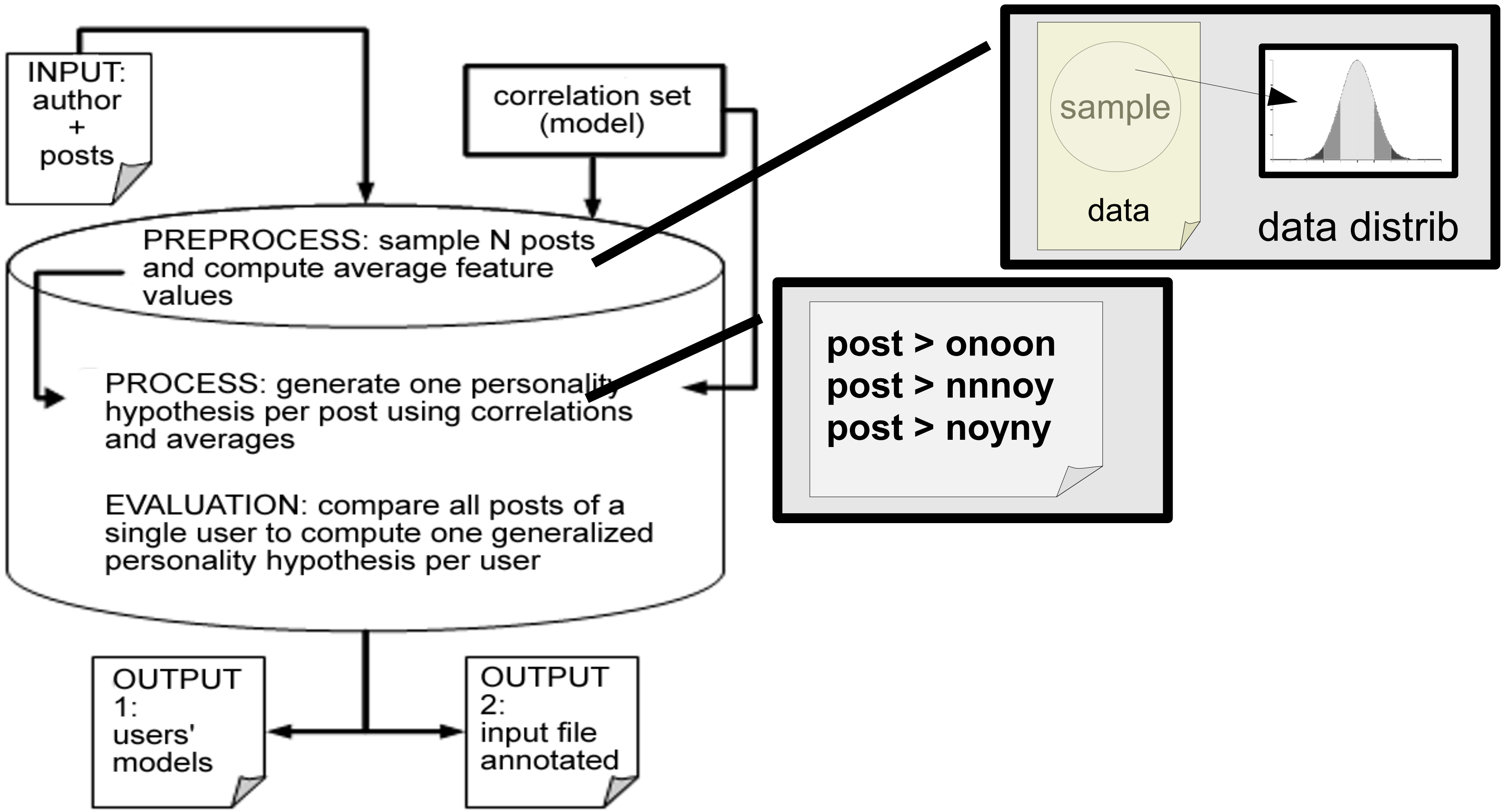
We need:

-unlabeled text + authors (many texts per author)
-small labeled test set
-correlations between language and personality

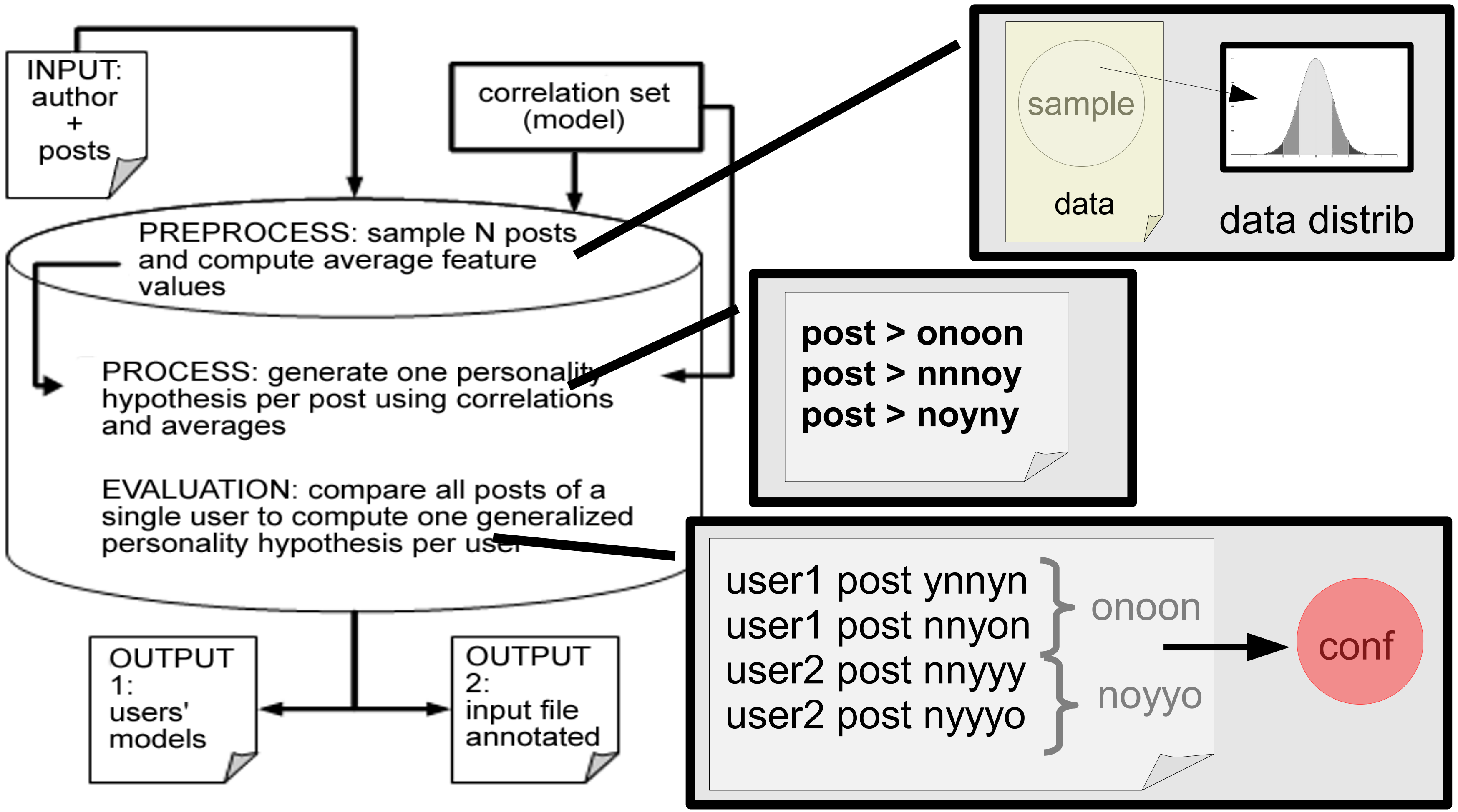
Fabio Celli
fabio.celli@unitn.it



Fabio Celli
fabio.celli@unitn.it

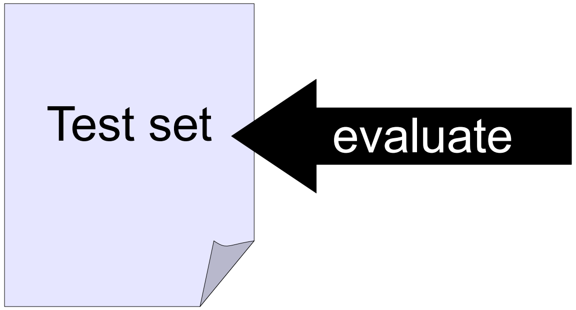


Fabio Celli
fabio.celli@unitn.it



Sep 16, 2014

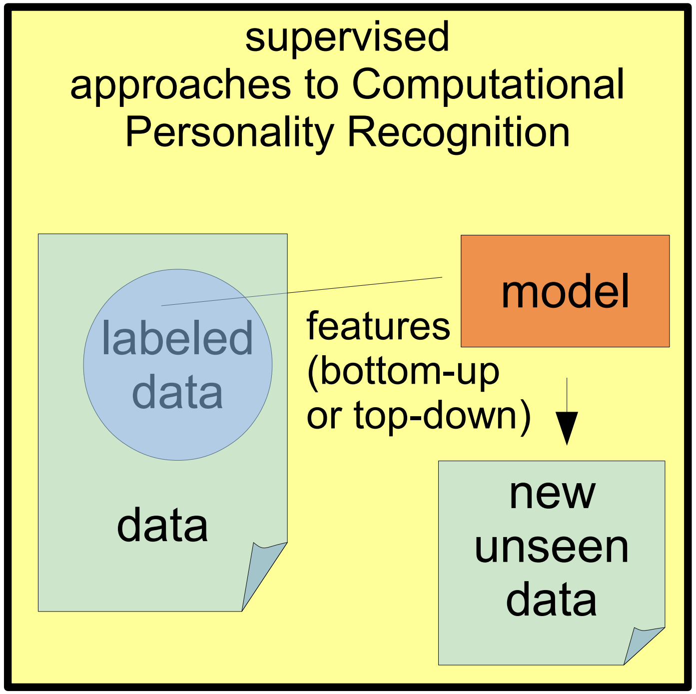
Fabio Celli
fabio.celli@unitn.it



onoon

noyyo

Fabio Celli
fabio.celli@unitn.it

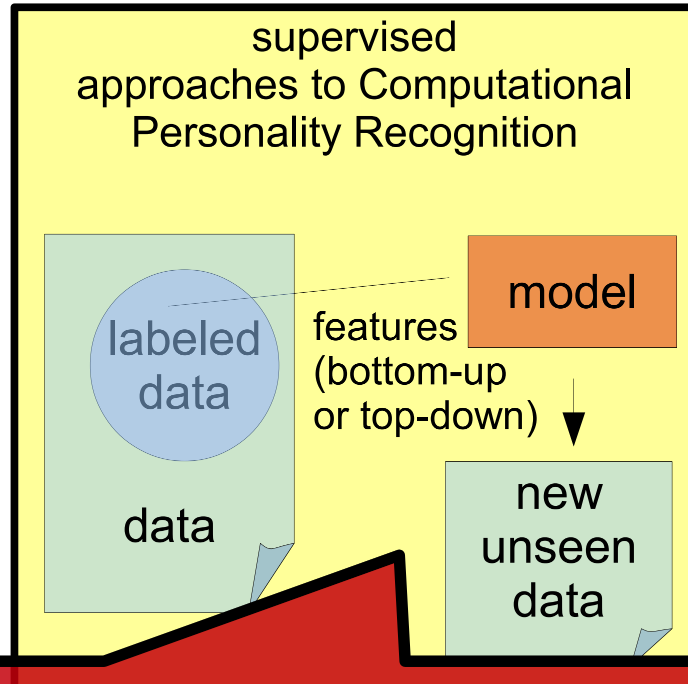


Problems of supervised:

1) <u>overfitting</u> → social network data samples are too small to extract good models and bottom up approaches extract very few good patterns

2) <u>multilinguality</u> → top down approaches use language dependent resources

Fabio Celli
fabio.celli@unitn.it

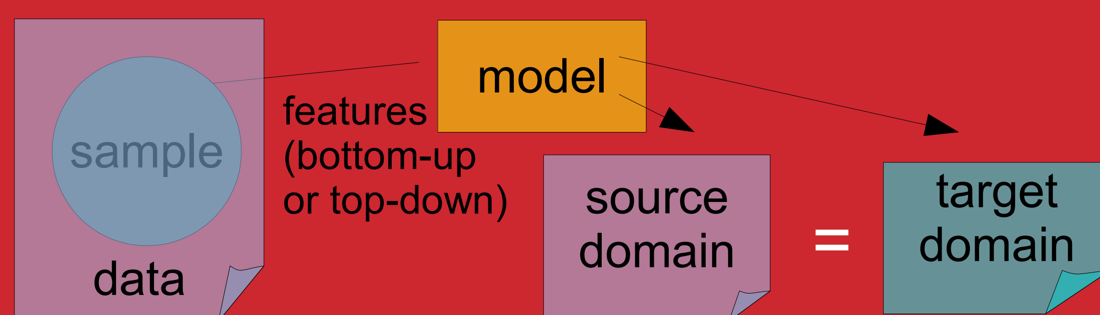


Problems of supervised:

1) <u>overfitting</u> → social network data samples are too small to extract good models and bottom up approaches extract very few good patterns

2) <u>multilinguality</u> → top down approaches use language dependent resources

Domain adaptation is a learning problem where a model is generalized across domains, and it is successfull when it minimizes the difference of performance from a source to a target domain [BenDavid et Al. 2006]



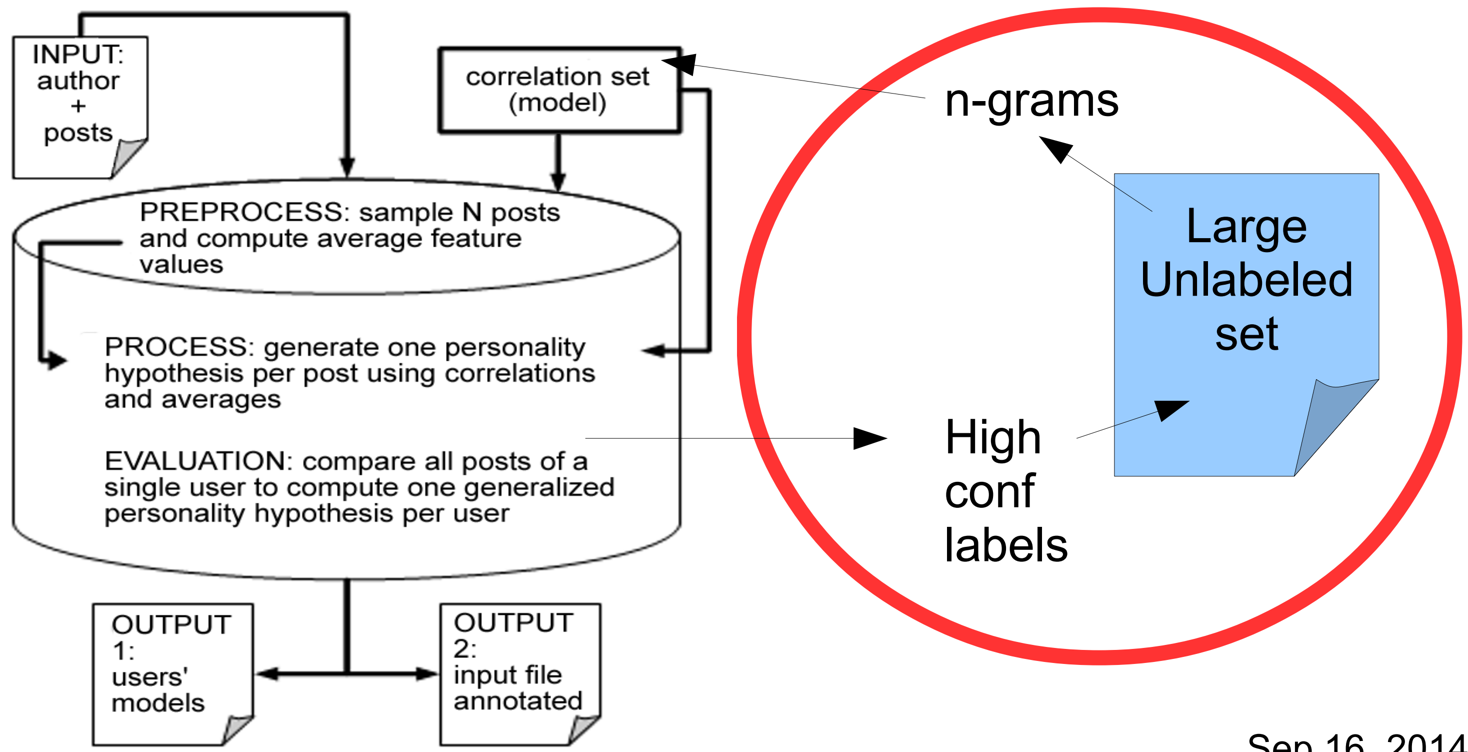
Avantage of unsupervised Personality recognition:

- <u>domain adaptability</u>

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

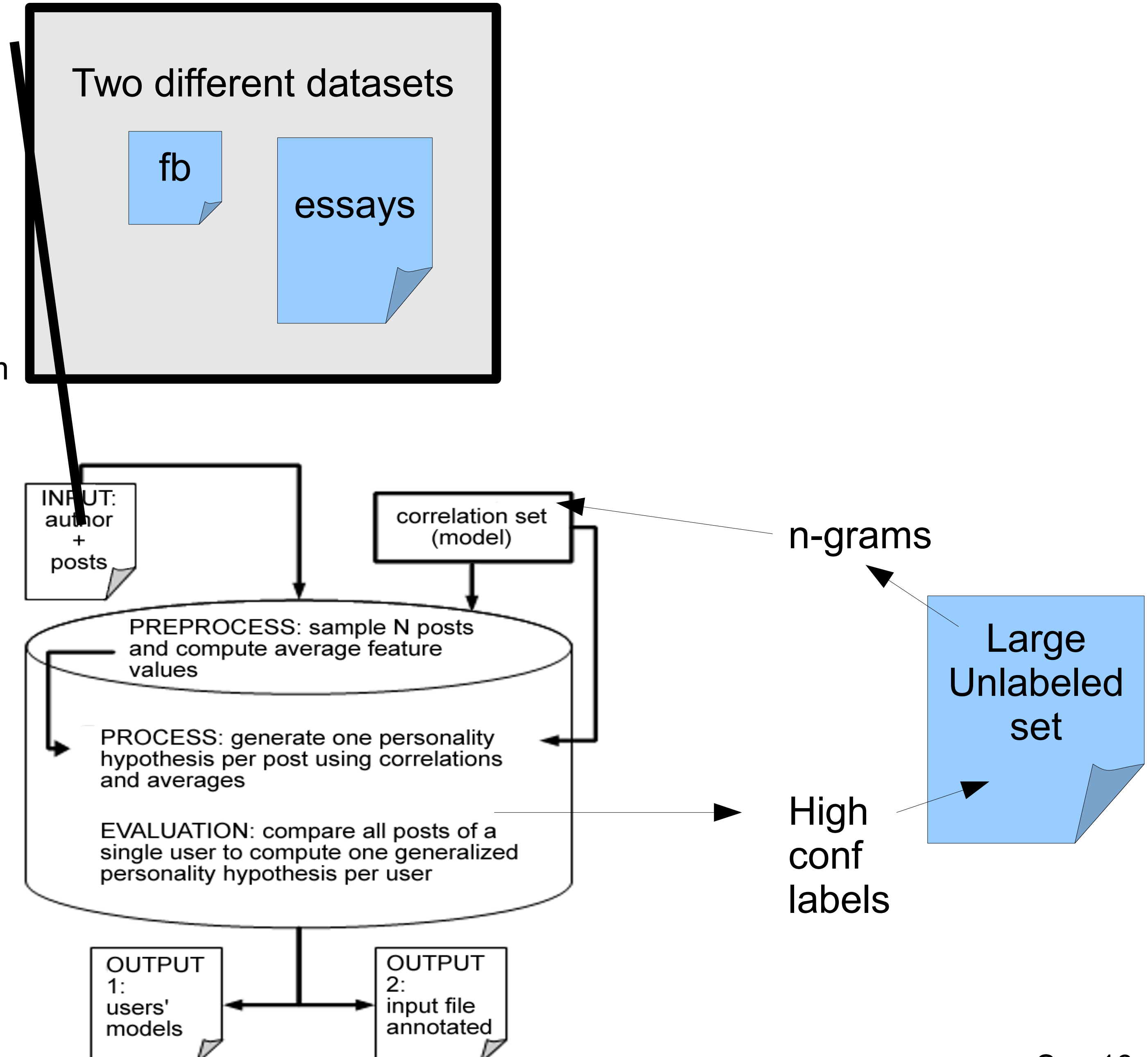
We added a part of the algorithm (semi-supervised). We explot the high confidence predictions from the unsupervised system to label an unlabeled large training set and extract n-grams from there that we add to the initial correlation set



Fabio Celli
fabio.celli@unitn.it

essays
[Pennebaker & King 1999]
[Mairesse et Al. 2007]
is a big collection of stream of consciousness writings of studentswho took the big5.
Lang: English
Unlabeled= ~2000 users
Test= ~200 users

PersFB
[Celli & Polonio (2013)]
is a small collection of Facebook statuses of students who took the big5.
Lang: Italian.
Unlabeled= ~200 users
Test= ~30 users



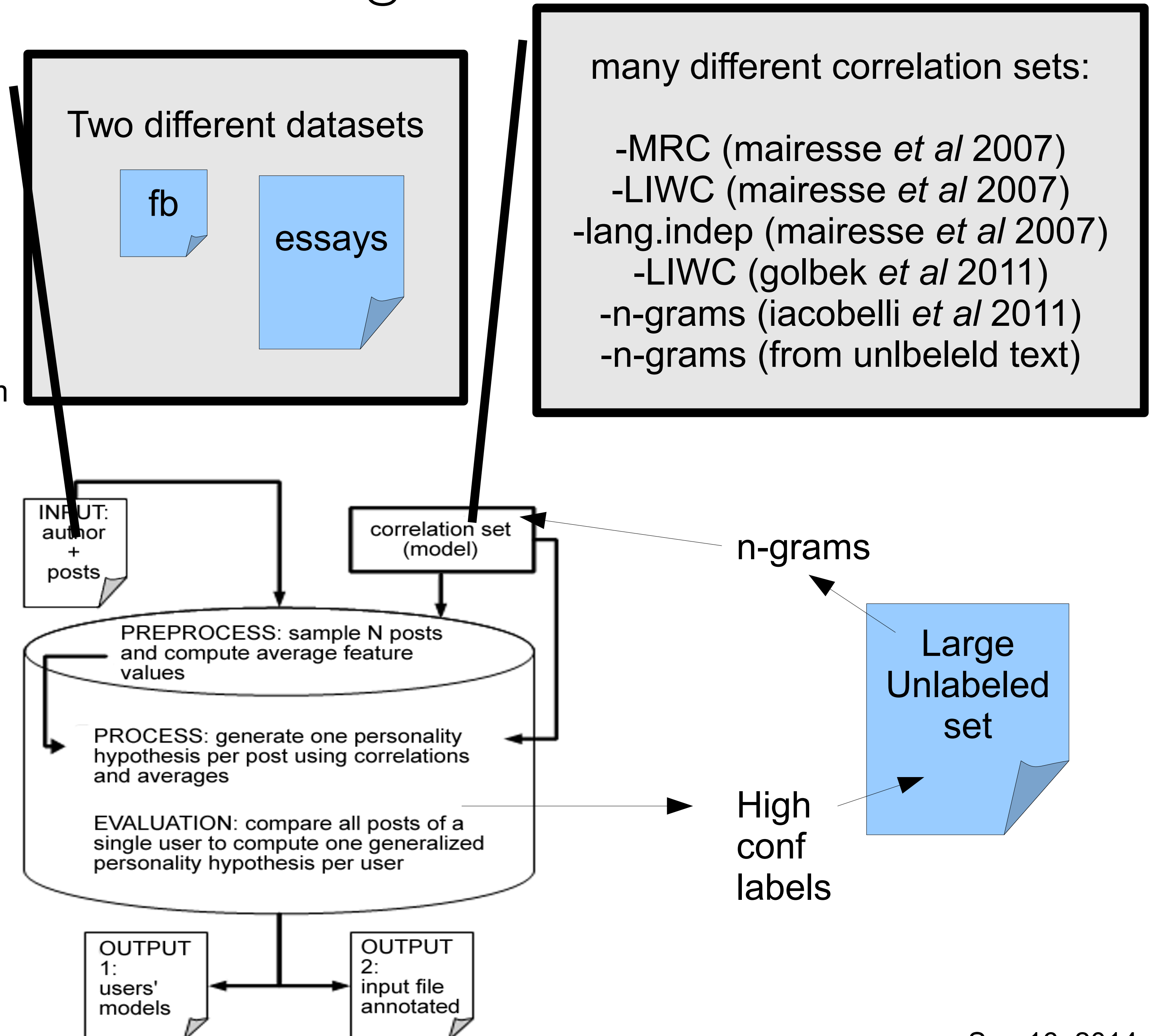
Fabio Celli
fabio.celli@unitn.it

essays

[Pennebaker & King 1999]
[Mairesse et Al. 2007]
is a big collection of stream of consciousness writings of studentswho took the big5.
Lang: English
Unlabeled= ~2000 users
Test= ~200 users

PersFB

[Celli & Polonio (2013)]
is a small collection of Facebook statuses of students who took the big5.
Lang: Italian.
Unlabeled= ~200 users
Test= ~30 users



Fabio Celli
fabio.celli@unitn.it

many different correlation sets:

-MRC (mairesse *et al* 2007)
-LIWC (mairesse *et al* 2007)
-lang.indep (mairesse *et al* 2007)
-LIWC (golbek *et al* 2011)
-n-grams (iacobelli *et al* 2011)
-n-grams (from unlbeleld text)

Fabio Celli
fabio.celli@unitn.it

12dimensions:

Nchar, Nphon, Nsyl,
Kffrq, Kfcat, Brownfrq
Tlfrq,
Conc, Fam
Imag, aoa

many different correlation sets:

-MRC (mairesse *et al* 2007)
-LIWC (mairesse *et al* 2007)
-lang.indep (mairesse *et al* 2007)
-LIWC (golbek *et al* 2011)
-n-grams (iacobelli *et al* 2011)
-n-grams (from unlbeleld text)

60+ dimensions

Posemo, negemo,
Anx, anger, sad,
Cogmech, insight, cause,
Certain, incl, excl
See, hear, feel,
Bio, body, health, sex,
Space, time, work
Achieve, leisure, home,
Money, relig, death
…
...

many different correlation sets:

-MRC (mairesse *et al* 2007)
-LIWC (mairesse *et al* 2007)
-lang.indep (mairesse *et al* 2007)
-LIWC (golbek *et al* 2011)
-n-grams (iacobelli *et al* 2011)
-n-grams (from unlbeleld text)

Fabio Celli
fabio.celli@unitn.it

many different correlation sets:

-MRC (mairesse *et al* 2007)
-LIWC (mairesse *et al* 2007)
-lang.indep (mairesse *et al* 2007)
-LIWC (golbek *et al* 2011)
-n-grams (iacobelli *et al* 2011)
-n-grams (from unlbeleld text)

feat	X	E	A	C	O
punctuation	-.08**	-.04	-.01	-.04	-10**
excl. marks	-.00	-.05*	.06**	.00	-.03
numbers	-.03	.05*	-.03	-.02	-.06**
parentheses	-.06**	.03	-.04*	-.01	.10**
quest. marks	-.06**	-.05*	-.04	-.06**	.08**
quotes	-.05*	-.02	-.01	-.03	.09**
repeat ratio	-.05**	.10**	-.04*	-.05*	.09**
avg word freq.	.05*	-.06**	.03*	.06**	.05**

Fabio Celli
fabio.celli@unitn.it

many different correlation sets:

-MRC (mairesse *et al* 2007)
-LIWC (mairesse *et al* 2007)
ng.indep (mairesse *et al* 2007)
IWC (golbek *et al* 2011)
s (iacobelli *et al* 2011)
ams (from unlbeleld text)

Trait	High	Low
N	hope.thei; punish.for; get.work; onli.problem; you.onli; depress.you; drunk.i; i.wasnt;	mental.togeth; be.sad; am.excit; we'v.had; reflect.on; then.look group.of; chose.to; the.winner
E	more.excit; i.hang; im.at; im.too; b**ch.i; danc.i; love.me; i.miss; you.f**k; wa.f**k; fun.anywai; hear.you; friend.were; love.me; a.club;	wai.so; my.regular; increas.my; my.flower; didn't.need; coupl.year; each.year; bond.slowli; favourit.charact; most.social; other.job;
O	is.beauti; like.s**t; be.held; think.he's; unabl.to; and.fun; danc.and; pick.me; i.lost; the.hell;	to.church; prai.for; at.church; laid.back; mondai.and; not.bad; you.belong; not.exactli; over.time;
A	even.better; of.beauti; compromis.with; hold.you; the.colleg; keep.myself; me.sigh; no.point; from.peopl;	like.it's; comment.about; like.it'; ex-cus.to; later.if; suppos.to; wa.worri; my.offic; sai.thing; goal.is; remain.in; return.of; send.the; unfortun.the; self.interest;
C	and.reliabl; prior.to; succe.in; so.hopefulli; got.caught; the.obviou; do.after; made.for; our.own; of.tear; on.track; to.drag; i.studi; hope.i'm; forget.that; realli.look;	episod.of; be.treat; not.thi; thi.just; pat-tern.is; real.reason; am.also; i.laugh; how.i'm; dare.to; of.why;

Fabio Celli
fabio.celli@unitn.it

resullts

Evaluation

Since each personality trait is bipolar, we considered:

true positives = correct predictions for both

false positives = wrong predictions for both

Fabio Celli
fabio.celli@unitn.it

resullts

Evaluation

Since each
personality trait
is bipolar,
we considered:

true positives =
correct predictions for both

false positives =
wrong predictions for both

Results

dataset	parameters	avg F1
persfb	rand baseline (2c)	.608
essays	rand baseline (2c)	.655
persfb	All features (2c)	**.686**
essays	All features (2c)	**.686**

=

Fabio Celli
fabio.celli@unitn.it

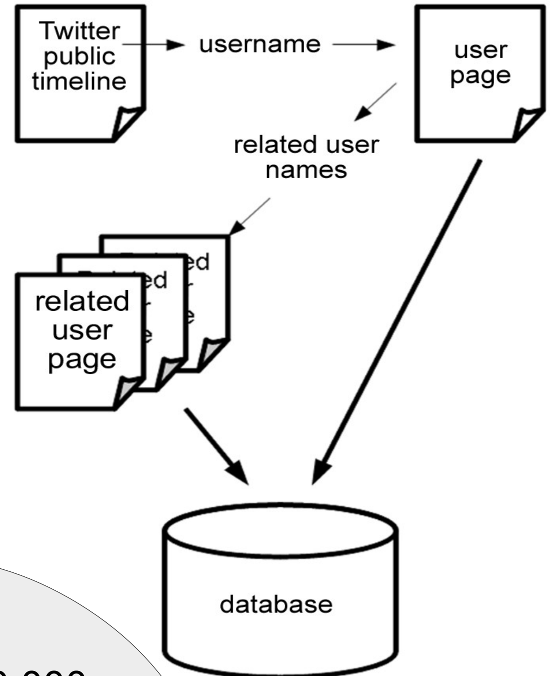
Applications of Unsupervised personality recognition from text

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

emotional stability
in Twitter
Conversations
[Celli & Rossi 2012]

Fabio Celli
fabio.celli@unitn.it

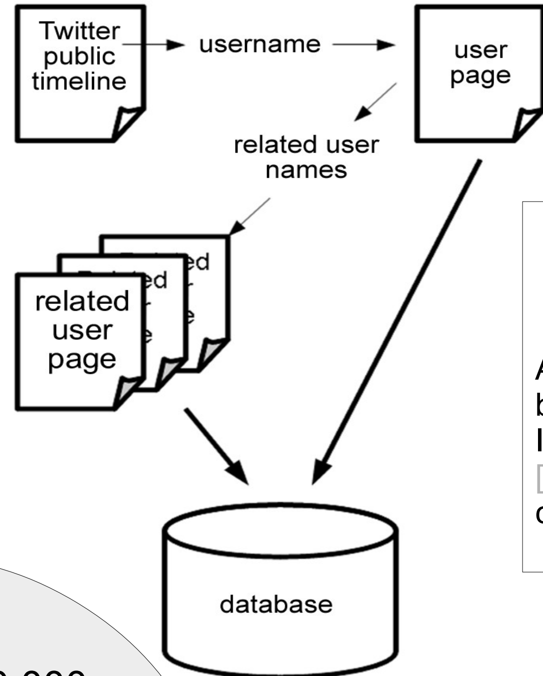


emotional stability in Twitter Conversations
[Celli & Rossi 2012]

collected > 200.000 posts and > 13.000 Authors. automatically annotated withPersonality (Secure / Neurotic)
+
added new correlations extracted from Twitter from recent literature
[Quercia et Al. 2011]

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it



emotional stability in Twitter Conversations

[Celli & Rossi 2012]

collected > 200.000 posts and > 13.000 Authors. automatically annotated withPersonality (Secure / Neurotic)
+
added new correlations extracted from Twitter from recent literature [Quercia et Al. 2011]

ANALYZE WORDS

AnalyzeWords helps reveal your personality by looking at how you use words in Twitter. It is based on good scientific research [Pennebaker et al 2001] connecting word use to who people are.

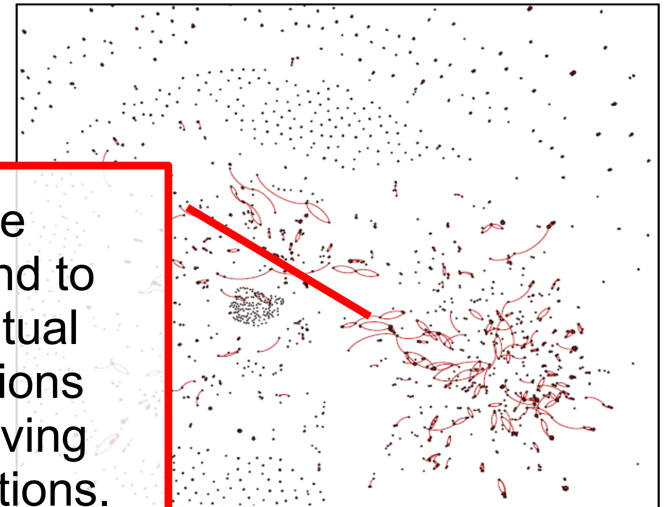
validation: comparison against analyzewords.com and essays

	p	r	f1
n	0.8	0.615	0.695
s	0.375	0.6	0.462
avg	0.587	0.607	0.578

Fabio Celli
fabio.celli@uni

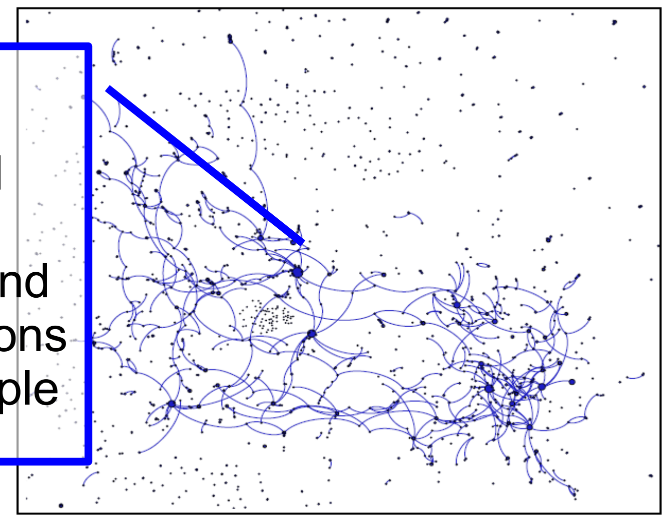
emotional stability in Twitter Conversations
[Celli & Rossi 2012]

collected > 200.000 posts and > 13.000 Authors. automatically annotated withPersonality (Secure / Neurotic)
+
added new correlations extracted from Twitter from recent literature
[Quercia et Al. 2011]



Secure users tend to build mutual connections while having conversations.

s

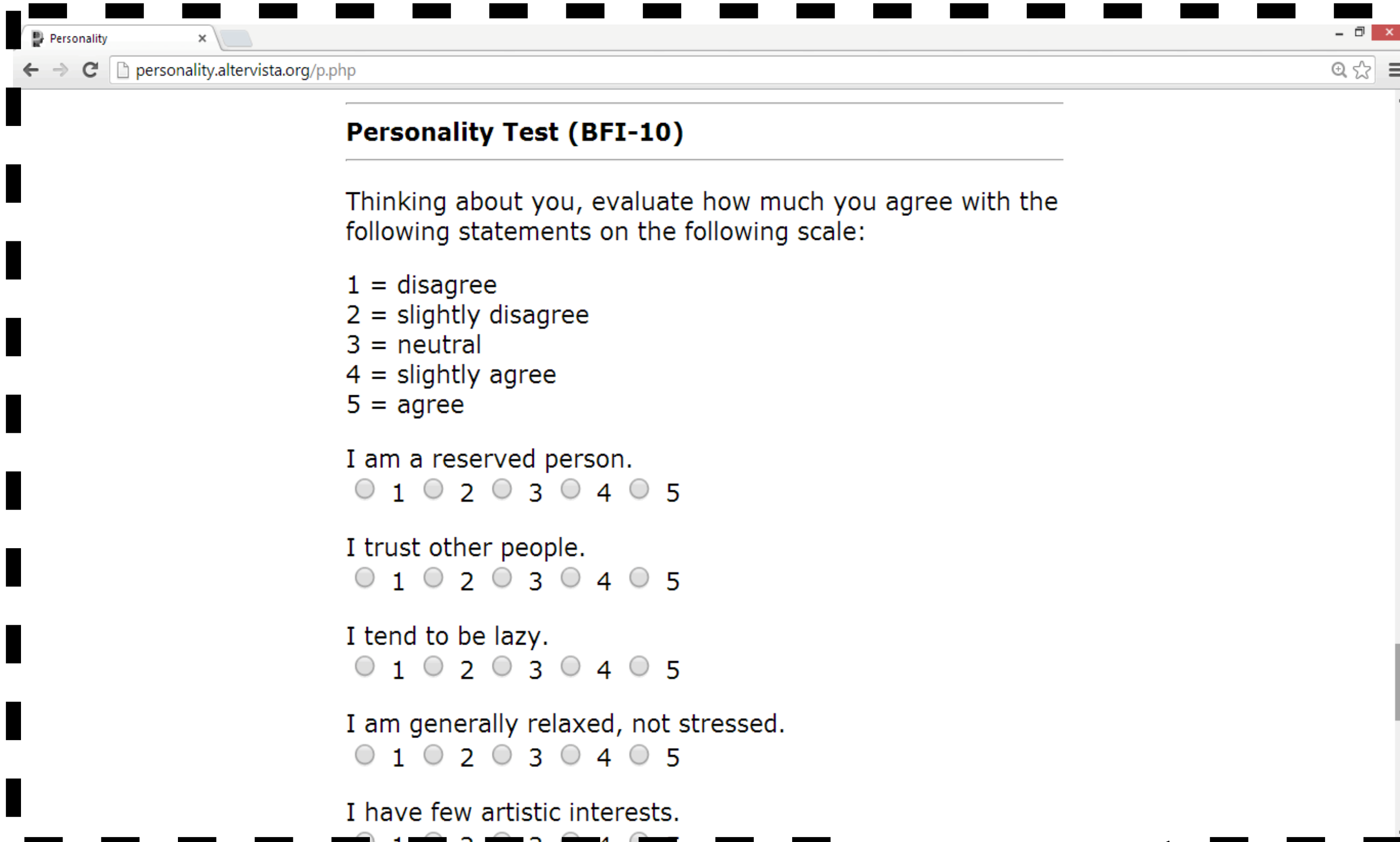


Neurotic users instead tend to build longer chains and have conversations with distant people

n

Fabio Celli
fabio.celli@unitn.it

emotional stability in Twitter Conversations
[Celli & Rossi 2012]



We are collecting the personality of Twitter users with 2 apps:

http://personality.altervista.org/personalitwit.php

(under dev)
http://personality.altervista.org/mypersonality/en/mypersonality.php

Fabio Celli
fabio.celli@unitn.it

Analysis of
Ego-Networks
in Facebook
[Celli & Polonio 2013]

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

Analysis of
Ego-Networks
in Facebook
[Celli & Polonio 2013]

collected > 5.000 posts and > 100 authors
from one access user, automatically annotated with Personality types

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

Analysis of Ego-Networks in Facebook
[Celli & Polonio 2013]

collected > 5.000 posts and > 100 authors from one access user, automatically annotated with Personality types

test set: 23 students took Big5 test and fb + off data

	p	r	f
off	.45	.7	.558
fb	.547	.735	.628

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

Analysis of
Ego-Networks
in Facebook
[Celli & Polonio 2013]

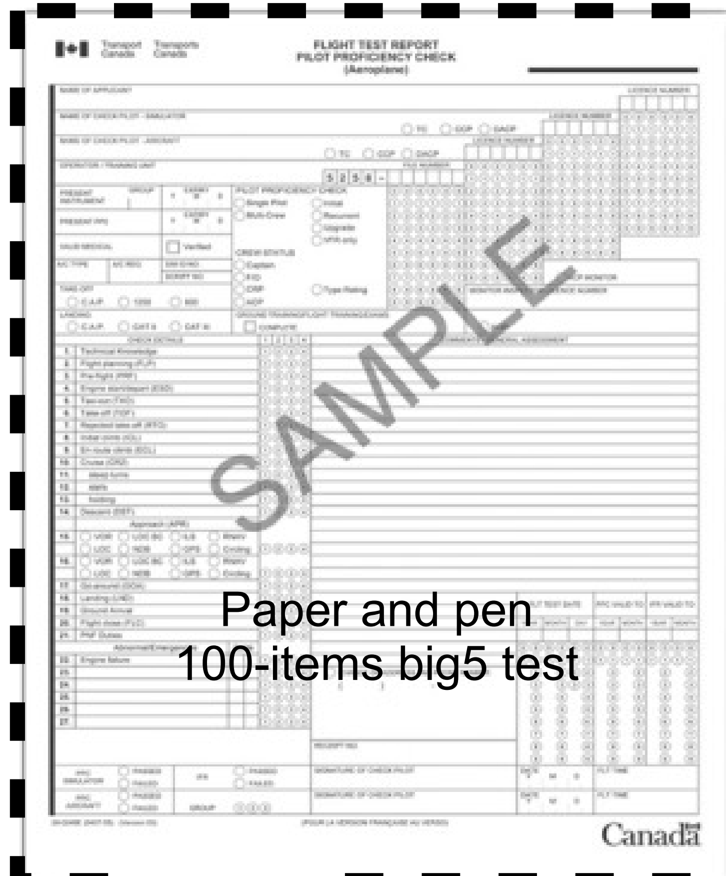
access-user's
timeline

seed users'
timeline

data

collected > 5.000 posts and > 100 authors from one access user, automatically annotated with Personality types

test set:
23 students took Big5 test and fb + off data



Paper and pen
100-items big5 test

	p	r	f
off	.45	.7	.558
fb	.547	.735	.628

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

Analysis of Ego-Networks in Facebook

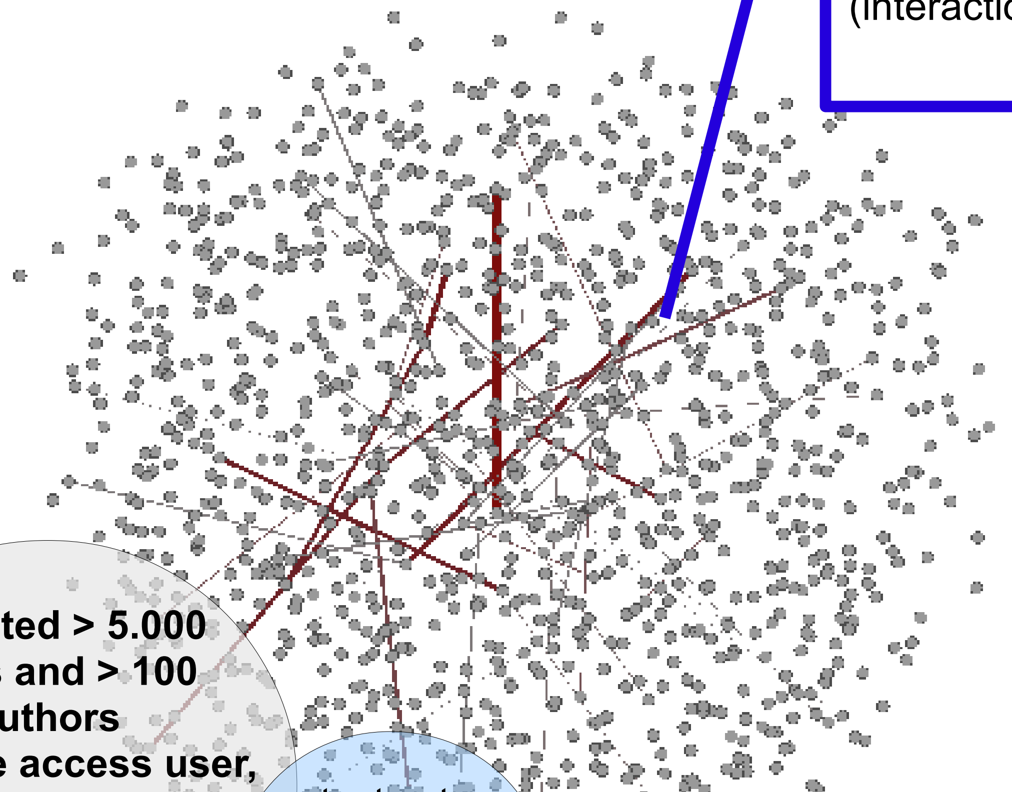
[Celli & Polonio 2013]

access-user's timeline

seed users' timeline

data

collected > 5.000 posts and > 100 authors from one access user, automatically annotated with Personality types



Open minded and introvert users have the highest Edge weight (interaction strength)

test set: 23 students took Big5 test and fb + off data

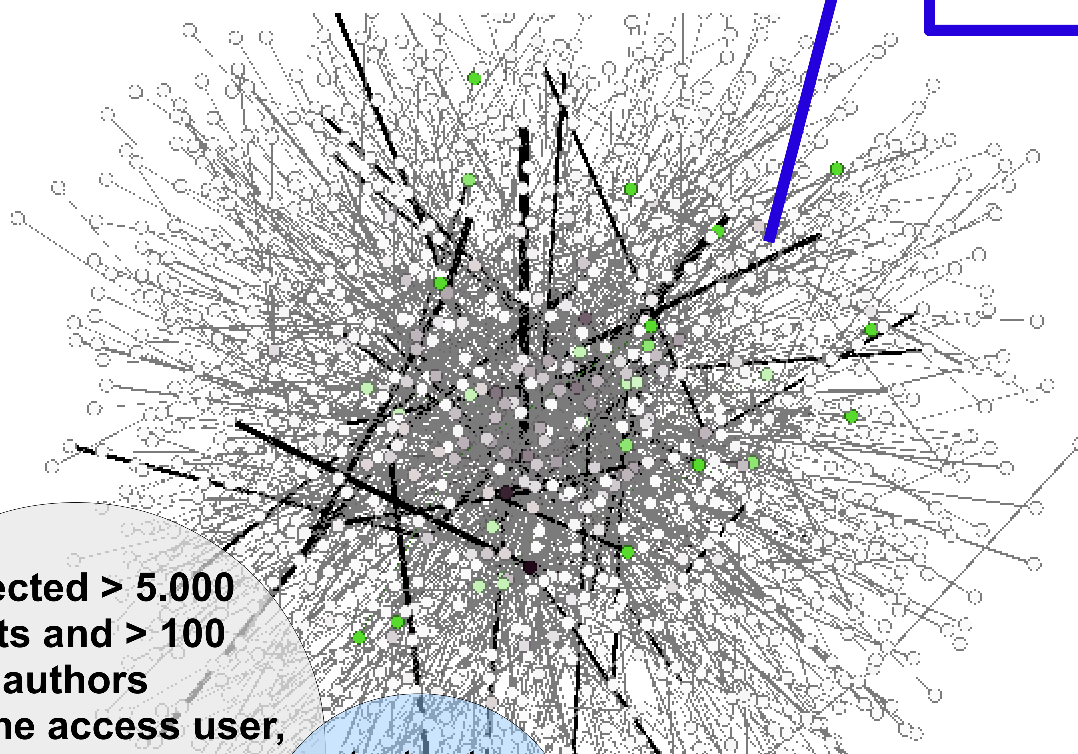
	p	r	f
off	.45	.7	.558
fb	.547	.735	.628

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

Uncooperative users have the highest clustering coefficient nodes that tend to participate to conversations

Analysis of Ego-Networks in Facebook
[Celli & Polonio 2013]



collected > 5.000 posts and > 100 authors from one access user, automatically annotated with Personality types

test set: 23 students took Big5 test and fb + off data

	p	r	f
off	.45	.7	.558
fb	.547	.735	.628

Sep 16, 2014

Fabio Celli
fabio.celli@unitn.it

Deception Detection
Via Personality
[Fornaciari et. al. 2013]

Fabio Celli
fabio.celli@unitn.it

Deception Detection
Via Personality
[Fornaciari et. al. 2013]

Can we detect liars exploiting personality?
Data: DeCour, 35 defendants from
4 hearings guity for calumny and false testimony
in 4 different Italian courts
Language: Italian

DeCour

Task: predict deceptions using
personality traits as features

Fabio Celli
fabio.celli@unitn.it

Deception Detection
Via Personality
[Fornaciari et. al. 2013]

Can we detect liars exploiting personality?
Data: DeCour, 35 defendants from 4 hearings guity for calumny and false testimony in 4 different Italian courts
Language: Italian

DeCour

Task: predict deceptions using personality traits as features

DECEPTION CLASSIFICATION VIA PERSONALITY

algorithm	P	R	F
mbl (zeroR)	0.313	0.56	0.402
dt (J4.8)	0.579	0.586	0.55
nb (NaiveBayes)	0.548	0.562	0.538
svm (SMO)	0.582	0.585	0.533
ripper (JRip)	0.576	0.582	0.532

averaged over the 5 traits

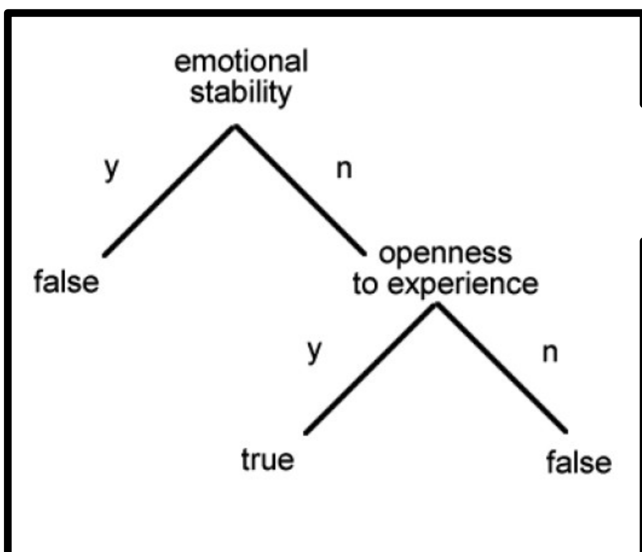
Fabio Celli
fabio.celli@unitn.it

Deception Detection
Via Personality
[Fornaciari et. al. 2013]

Can we detect liars exploiting personality?
Data: DeCour, 35 defendants from
4 hearings guity for calumny and false testimony
in 4 different Italian courts
Language: Italian

DeCour

Task: predict deceptions using
personality traits as features



DECEPTION CLASSIFICATION VIA PERSONALITY

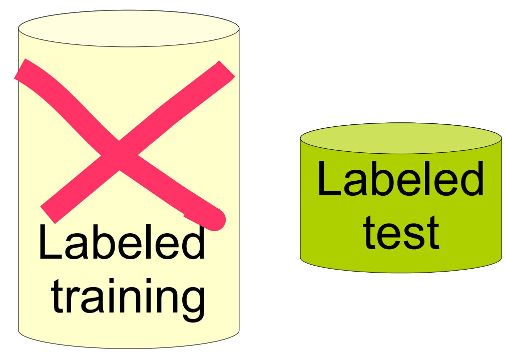
algorithm	P	R	F
mbl (zeroR)	0.313	0.56	0.402
dt (J4.8)	0.579	0.586	0.55
nb (NaiveBayes)	0.548	0.562	0.538
svm (SMO)	0.582	0.585	0.533
ripper (JRip)	0.576	0.582	0.532

averaged over the 5 traits

Fabio Celli
fabio.celli@unitn.it

Summing up:

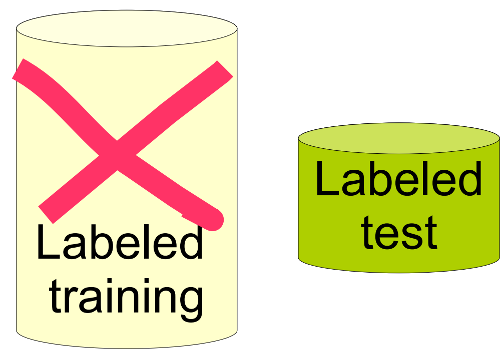
Unsup./semisup. Personality recognition is useful in those domains where it is difficult to retrieve labeled data



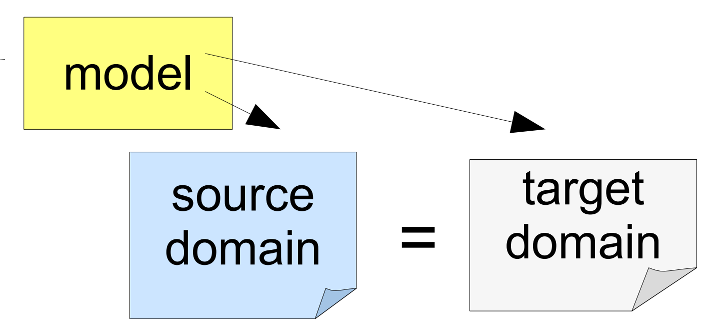
Fabio Celli
fabio.celli@unitn.it

Summing up:

- Unsup./semisup. Personality recognition is useful in those domains where it is difficult to retrieve labeled data



- It is domain adaptive



Fabio Celli
fabio.celli@unitn.it

in conclusion:

-supervised:
domain dependent,
high performance

-unsupervised:
adaptability,
applicability in extreme conditions


thank you
рахмат
danke
謝謝
ngiyabonga
teşekkür ederim
Баярлалаа
спасибо
gracias
merci
obrigado
dziękuję
bedankt
go raibh maith agat
mochchakkeram
sukriya
kop khun krap
grazie
arigatô
dakujem
terima kasih
감사합니다
ありがとう
мерси
tapadh leat
dank je
hvala
mauruuru
спасибі
хвала
дякую
ευχαριστώ