Introduction to Machine Learning
NPFL 054

http://ufal.mff.cuni.cz/course/npfl054

Barbora Hladká **Martin Holub**

{Hladka | Holub}@ufal.mff.cuni.cz

Charles University,
Faculty of Mathematics and Physics,
Institute of Formal and Applied Linguistics

Demo 1
Verb Patterns Classification

Purpose of the demo task
= to show several things related to gold standard data for a supervised machine learning task, especially

- Manual annotation and basic data analysis
- Gold Standard data distribution
- Inter-annotator agreement
- Confusion matrices
- Error analysis

Verb Patterns Classification – task description

Verb Patterns Classification is a kind of *lexical disambiguation* of verbs. The task is similar to the traditional *word sense disambiguation* (WSD). The two tasks differ in how the semantic categories are defined (word senses vs. patterns of typical verb usage).

Let's focus on two English verbs, namely *cry* and *enlarge*.

CRY -- dictionary definitions

cry ◆◆◆◆◇

1 cry; cries; crying; cried
When you cry, tears come from your eyes, usually because you are unhappy or hurt.
I hung up the phone and started to cry.
Please don't cry.
He cried with anger and frustration.
...a crying baby.
VB

2 cry; cries; crying; cried
If you cry something, you shout it or say it loudly.
`Nancy Drew,' she cried, `you're under arrest!'.
I cried: `It's wonderful news!'
VB

5 cry; cries
You can refer to a public protest about something or appeal for something as a cry of some kind. (JOURNALISM)
There have been cries of outrage about this expenditure.
Many other countries have turned a deaf ear to their cries for help.
N-COUNT: usu N of/for n

ENLARGE -- dictionary definitions

enlarge ◆◆◇◇◇ enlarge; enlarges; enlarging; enlarged

1 When you enlarge something or when it enlarges, it becomes bigger.
...the plan to enlarge Ewood Park into a 30,000 all-seater stadium...
The glands in the neck may enlarge.
V-ERG

@ enlarged
The UN secretary-general yesterday recommended an enlarged peacekeeping force.
ADJ-GRADED

2 To enlarge a photograph means to develop a bigger print of it.
...newly-weds wishing to enlarge snaps of their big day.
VB

3 If you enlarge on something that has been mentioned, you give more details about it. (FORMAL)
He didn't enlarge on the form that the interim government and assembly would take.
I wish to enlarge upon a statement made by Gary Docking.
VB
= expand

CRY -- Pattern definitions

Pattern 1 **[Human] cry [no object]**

Explanation [[Human]] weeps
usually because [[Human]] is unhappy or in pain

Example *His advice to stressful women was: ` If you **cry**, do n't cry alone.*

Pattern 4 **[Human] cry [THAT-CL|WH-CL|QUOTE] ({out})**

Explanation [[Human]] shouts ([QUOTE]) loudly
typically, in order to attract attention

Example *You can hear them screaming and banging their heads, **crying** that they want to go home.*

Pattern 7 **[Entity | State] cry [{out}] [{for} Action] [no object]**

Explanation [[Entity | State]] requires [[Action]] to be taken urgently

Example *Identifying areas which **cry** out for improvement or even simply areas of muddle and misunderstanding, is by no means negative -- rather a spur to action.*

ENLARGE -- Pattern definitions

Pattern 1 **[[Human]^[Eventuality]] enlarge [Entity]**

Explanation [[Human | Eventuality]] causes [[Entity]] to grow or become larger

Example *These were not large powers, but later changes were to **enlarge** them.*

Pattern 2 **[Entity] enlarge [no object]**

Explanation [[Entity]] grows or becomes larger

Example *As infants grow, their bodies not only **enlarge** but change both in shape and colour.*

Pattern 3 **[[Human]^[Document]] enlarge [{on | upon} Anything = Topic] [no object]**

Explanation [[Human]] speaks or writes at length on [[Anything = Topic]] or [[Document]] contains long-winsed comments on [[Topic]]}

Example *Let me **enlarge** on this a little.*

Pattern 4 **enlarged**

Explanation now larger than before,without any deliberate causer or causer irrelevant

Example *The fluid filled spaces or ventricles appear to be **enlarged**, and the blood flow to the front of the brain is reduced.*

Verb Patterns Classification – annotation description

You will classify *cry* and *enlarge* manually.

- You will be given 10+10 sentences with the given verbs
- For each sentence you will assign a pattern that fits best the given sentence
  - there are 3 predefined patterns for the verb *cry*
  - there are 4 predefined patterns for the verb *enlarge*
  - if you think that no pattern matches the sentence, choose "u"
  - if you think that the given word is not a verb, choose "x"
- Use the forms posted at `https://ufal.mff.cuni.cz/courses/npfl054/demo`

Gold Standard data – distributions

Gold standard data sets are posted on the course web page (DEMO).

CRY – 250 instances in the GS set

class	1	4	7	u	x
frequency	131	59	13	33	14

ENLARGE – 300 instances in the GS set

class	1	2	3	4	u
frequency	230	21	20	26	3

Automatic classifier

Automatic classifier is a function that assigns certain output class to each input instance.

Output class is a discrete (possibly categorical) value.

In the demo task: Pattern tags are categorical output values, sentences containing the verbs in question are input instances.

Classifier accuracy is often *estimated* using a test data sample as a percentage of correctly classified instances in the sample. This estimate is called *sample accuracy*.

Automatic predictions made by automatic classifier (our best model F1) are posted on the course web page (DEMO).

- NOTE that it is the same GS set, and it was also used as training data (!).
- Thus, you can compute only the training error, not the test error.

Manual annotation

Annotated data – a subset of the GS

– the same data set annotated by each group

2014 – 2 groups

- **A** (5 Czech)
- **B** (2 Czech, 3 foreign)

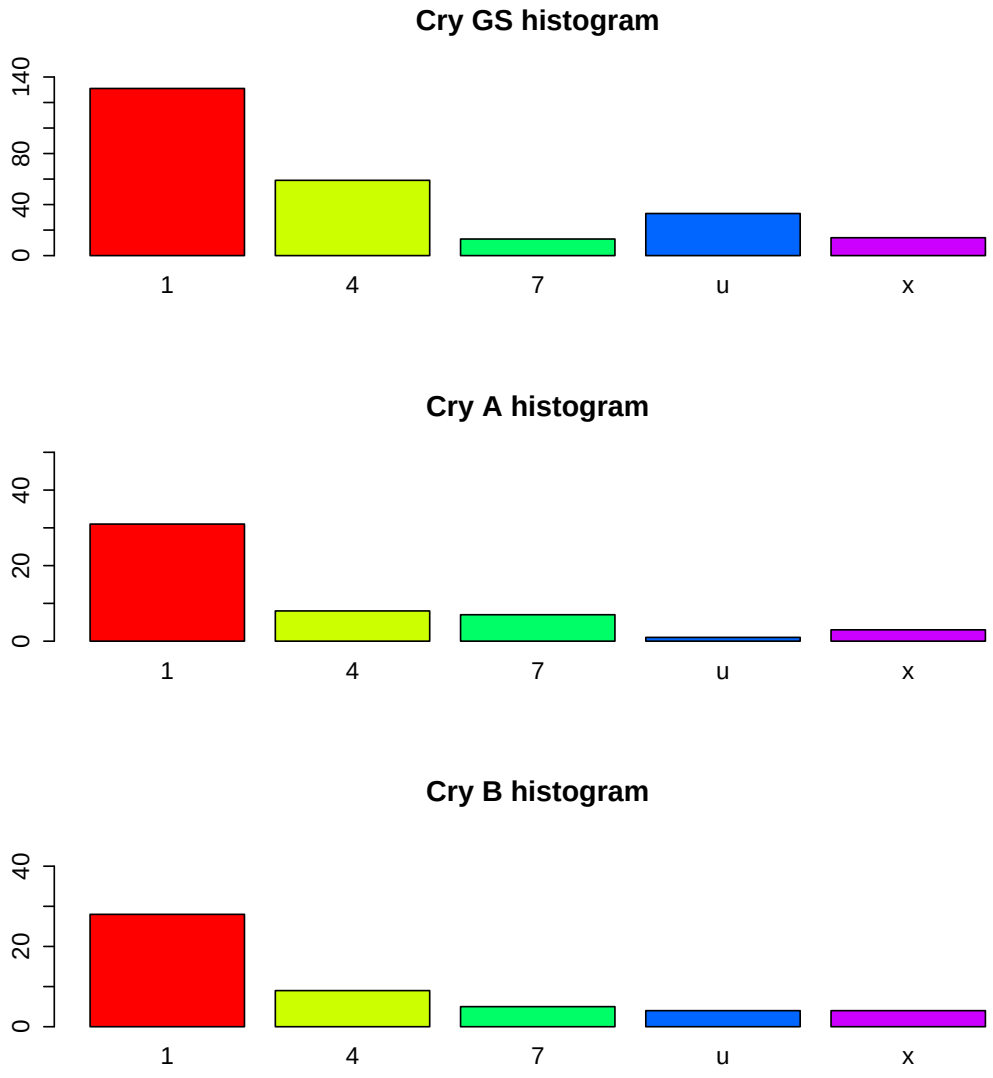
2015 – 4 groups

- **A** (6 Czech)
- **B** (6 Czech)
- **C** (6 Czech)
- **D** (6 Czech)

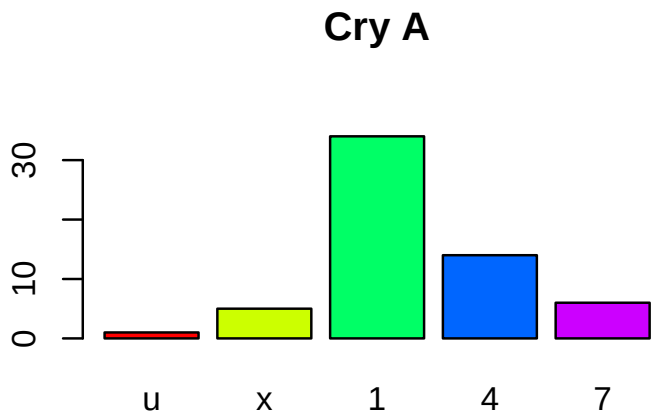
Now we can analyse/compare

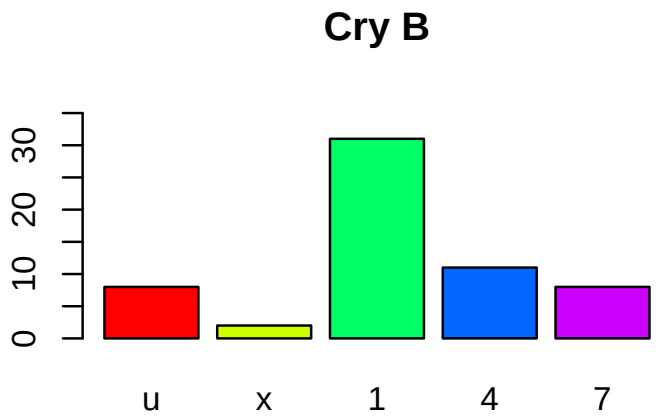
- which group is closer to the Gold Standard
- inter-annotator agreement between groups
- error types
  - – made by people
  - – made by automatic classifier

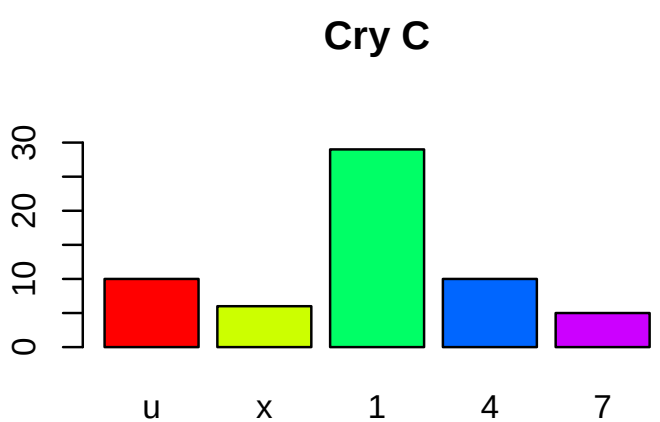
A, B and GS distributions - CRY (2014)

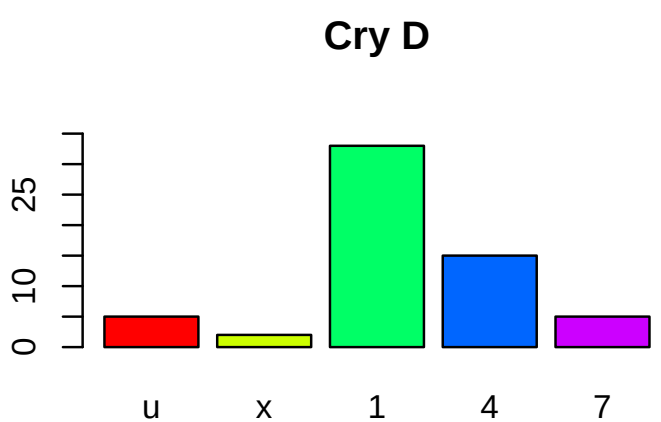


A, B, C, D distributions - CRY (2015)

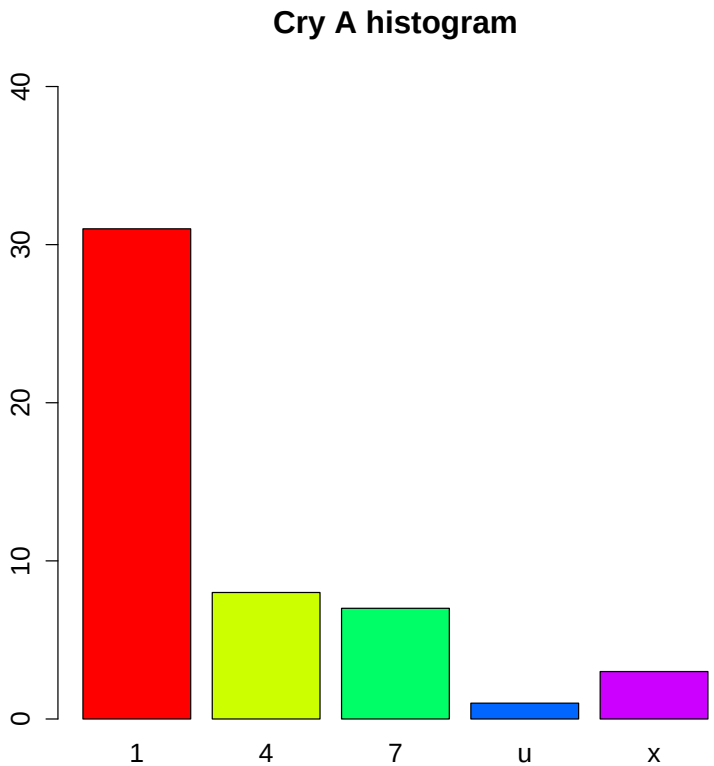








A vs GS - confusion matrix - CRY (2014)



		GS				
		1	**4**	**7**	**u**	**x**
	1	27	2	0	2	0
	4	1	6	0	0	1
A	**7**	0	1	2	3	1
	u	1	0	0	0	0
	x	2	0	0	1	0

Number of agreements: 35 (70 %)
Number of disagreements: 15 (30 %)

A, B, C, D vs GS - confusion matrix - CRY (2015)

A \ GS	u	x	1	4	7
u	0	0	1	0	0
x	0	2	2	1	0
1	2	0	29	3	0
4	1	0	5	8	0
7	3	0	1	0	2

Agreement: 41 (68 %)
Disagreement: 19 (32 %)

C \ GS	u	x	1	4	7
u	3	0	5	2	0
x	1	1	2	2	0
1	1	0	28	0	0
4	1	0	3	6	0
7	0	1	0	2	2

Agreement: 40 (67 %)
Disagreement: 20 (33 %)

B \ GS	u	x	1	4	7
u	2	1	3	2	0
x	0	0	2	0	0
1	0	0	30	1	0
4	2	0	3	6	0
7	2	1	0	3	2

Agreement: 40 (67 %)
Disagreement: 20 (33 %)

D \ GS	u	x	1	4	7
u	2	0	2	1	0
x	1	0	1	0	0
1	1	1	31	0	0
4	0	1	4	10	0
7	2	0	0	1	2

Agreement: 45 (75 %)
Disagreement: 15 (25 %)

A, B, C, D vs GS - confusion m. - ENLARGE (2015)

A \ GS	u	1	2	3	4
u	0	4	0	0	0
1	0	26	0	0	0
2	1	4	7	0	1
3	0	1	0	0	1
4	0	9	1	0	5

Agreement: 38 (63 %)
Disagreement: 22 (37 %)

C \ GS	u	1	2	3	4
u	0	5	2	0	1
1	1	18	0	0	0
2	0	6	6	0	2
3	0	4	0	0	0
4	0	11	0	0	4

Agreement: 28 (47 %)
Disagreement: 32 (53 %)

B \ GS	u	1	2	3	4
u	0	5	0	0	0
1	1	18	1	0	1
2	0	6	7	0	2
3	0	2	0	0	1
4	0	13	0	0	3

Agreement: 28 (47 %)
Disagreement: 32 (53 %)

D \ GS	u	1	2	3	4
u	0	4	0	0	0
1	1	25	2	0	1
2	0	3	6	0	1
3	0	2	0	0	0
4	0	10	0	0	5

Agreement: 36 (60 %)
Disagreement: 24 (40 %)

Inter-annotator agreement (IAA) (2014)

CRY – confusion matrix (50 instances, 33 agreements = 66 %)

		B				
		1	**4**	**7**	**u**	**x**
	1	24	3	1	3	0
	4	3	3	0	1	1
A	**7**	0	2	4	0	1
	u	1	0	0	0	0
	x	0	1	0	0	2

ENLARGE – confusion matrix (50 instances, 31 agreements = 62 %)

		B				
		1	**2**	**3**	**4**	**u**
	1	18	2	0	2	0
	2	4	7	1	4	0
A	**3**	0	0	0	0	0
	4	2	1	2	5	0
	u	0	0	0	1	1

What agreement would be reached by chance?

Example 1

Assume two annotators (A_1, A_2), two classes (t_1, t_2), and the following distribution:

	t_1	t_2
A_1	50 %	50 %
A_2	50 %	50 %

Then

- the best possible agreement is 100 %
- the worst possible agreement is 0 %
- the "agreement-by-chance" *would be* 50 %

What agreement would be reached by chance?

Example 2

Assume two annotators (A_1, A_2), two classes (t_1, t_2), and the following distribution:

	t_1	t_2
A_1	90 %	10 %
A_2	90 %	10 %

Then

- the best possible agreement is 100 %
- the worst possible agreement is 80 %
- the "agreement-by-chance" *would be* 82 %

What agreement would be reached by chance?

Example 3

Assume two annotators (A_1, A_2), two classes (t_1, t_2), and the following distribution:

	t_1	t_2
A_1	90 %	10 %
A_2	80 %	20 %

Then

- the best possible agreement is 90 %
- the worst possible agreement is 70 %
- the "agreement-by-chance" *would be* 74 %

Example in R

The situation from Example 3 can be simulated in R

```
# N will be the sample size
> N = 10^6

# two annotators will annotate randomly
> A1 = sample(c(rep(1, 0.9*N), rep(0, 0.1*N)))
> A2 = sample(c(rep(1, 0.8*N), rep(0, 0.2*N)))

# percentage of their observed agreement
> mean(A1 == A2)
[1] 0.740112

# exact calculation -- just for comparison
> 0.9*0.8 + 0.1*0.2
[1] 0.74
```

Cohen's kappa

Cohen's kappa was introduced by Jacob Cohen in 1960.

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

- $\Pr(a)$ is the relative observed agreement among annotators
  = percentage of agreements in the sample
- $\Pr(e)$ is the hypothetical probability of chance agreement
  = probability of their agreement if they annotated randomly
- $\kappa > 0$ if the observed agreement is better than what would be expected by chance

Limitations

- Cohen's kappa measures agreement between two annotators only
- for more annotators you should use the more general Fleiss' kappa
  – see `http://en.wikipedia.org/wiki/Fleiss'_kappa`

Inter-annotator agreement (2014)

CRY
Number of agreements: 33 (66 %)
Number of disagreements: 17 (34 %)
Cohen's kappa: 0.437
Fleiss's kappa: 0.434

ENLARGE
Number of agreements: 31 (62 %)
Number of disagreements: 19 (38 %)
Cohen's kappa: 0.438
Fleiss's kappa: 0.433

Inter-annotator agreement (2015)

CRY – **Cohen's kappa**

	A	B	C	D
A	–	0.36	0.28	0.41
B	–	–	0.37	0.41
C	–	–	–	0.33
D	–	–	–	–

ENLARGE – **Cohen's kappa**

	A	B	C	D
A	–	0.31	0.41	0.30
B	–	–	0.22	0.32
C	–	–	–	0.37
D	–	–	–	–

CRY – **Fleiss's kappa** 0.35
ENLARGE – **Fleiss's kappa** 0.32

Automatic classifier – training error analysis
ENLARGE (2014)

		GS				
		1	**2**	**3**	**4**	**u**
C	**1**	224	1	1	12	2
	2	2	17	3	0	0
	3	1	2	15	0	0
	4	3	1	0	14	1
	u	0	0	1	0	0

	GS				
	1	**2**	**3**	**4**	**u**
1	0.97	0.05	0.05	0.46	0.67
2	0.01	0.81	0.15	0.00	0.00
3	0.00	0.10	0.75	0.00	0.00
4	0.01	0.05	0.00	0.54	0.33
u	0.00	0.00	0.05	0.00	0.00

Number of agreements: 270 (90 %)
Number of disagreements: 30 (10 %)

A + B error analysis – ENLARGE (2014)

		GS				
		1	**2**	**3**	**4**	**u**
A+B	**1**	46	0	0	0	0
	2	11	14	0	1	0
	3	3	0	0	0	0
	4	12	0	0	10	0
	u	0	0	0	1	2

	GS				
	1	**2**	**3**	**4**	**u**
1	0.64	0.00	0.00	0.00	0.00
2	0.15	1.00	0.00	0.08	0.00
3	0.04	0.00	0.00	0.00	0.00
4	0.17	0.00	0.00	0.83	0.00
u	0.00	0.00	0.00	0.08	1.00

Number of agreements: 72 (72 %)
Number of disagreements: 28 (28 %)

Summary of Lab #1
Examination Requirements

You should be able to practically compute and understand/use

- categorical data distribution
- confusion matrices
- classifier accuracy
- inter-annotator agreement
  - simple percentage
  - Cohen's kappa
- probability (both conditional and unconditional) of errors of different types

Practical exercises in R

- Download two files with annotated data cry-A.csv and cry-C.csv.
  - `https://ufal.mff.cuni.cz/courses/npfl054/demo`
- Run R and read the data using `read.csv()`.
  - Hint: see the posted Tutorial, Part I.
  - ... and create objects `cry.A` and `cry.C`.
- Make the confusion matrix between groups A and C.
  - Hint: use `table(cry.A$class, cry.C$class)`
- Compute simple agreement (in percentage) between A and C.
  - Hint: use `diag()` and `sum()`
- compute the Cohen's kappa value between groups A and C.
  - For hints see Part III of the Tutorial.

Summary of Lab #1
Homework

- Go through all details in the Tutorial (Parts I, II, and III)
- Get familiar with the `data.table` package
  – just to understand Part II
- Do all exercises in Part III