

---

---

# Sentence-level Emotion Classification with Label and Context Dependence

— *Daniela Ryšavá* —

---

---

# Emotion Classification

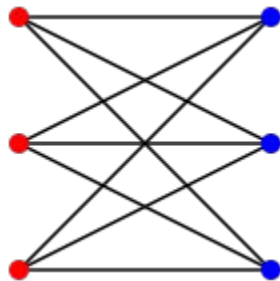
- predikce emocí v textu (*hněv, radost, láska, překvapení, smutek, ...*)
- výzkum na úrovni dokumentu → věta
- využití př.: analýza textu na sociálních sítích

# Sentence-level Emotion Classification

- klasifikace emocí na úrovni věty
- problémy:
  - klasifikace multi-label („You're the best wife ever!“ – *radost, láska*)
  - nedostatek informací („That is all I want.“ – *radost?*)
- Dependence Factor Graph (DFG) = nový přístup přihlížející k:
  - závislosti labelů (*láska + radost, radost × smutek*)
  - závislost na kontextu

# Dependence Factor Graph (DFG)

- modelování závislosti labelů a kontextu v klasifikaci emocí na úrovni věty (první přístup zahrnující obě závislosti v jednom rámci)
- = bipartitní graf, představuje rozklad funkcí (1 závislost = 1 fce)



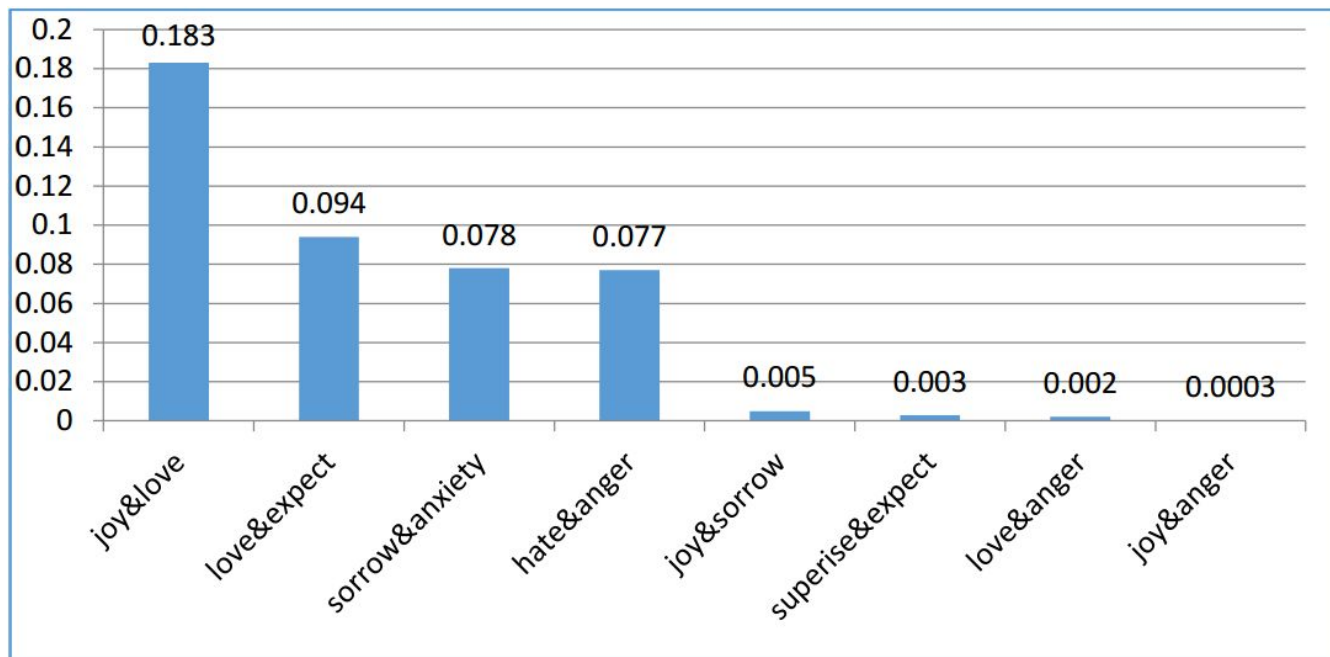
# Korpusová data: závislost labelů

- 100 náhodných dokumentů, 2751 označovaných vět
- pravděpodobnost, že se 2 emoce vyskytnou současně

Emotion	#Sentence	Emotion	#Sentence
<i>joy</i>	691	<i>anxiety</i>	567
<i>hate</i>	532	<i>surprise</i>	180
<i>love</i>	1025	<i>anger</i>	287
<i>sorrow</i>	611	<i>expect</i>	603

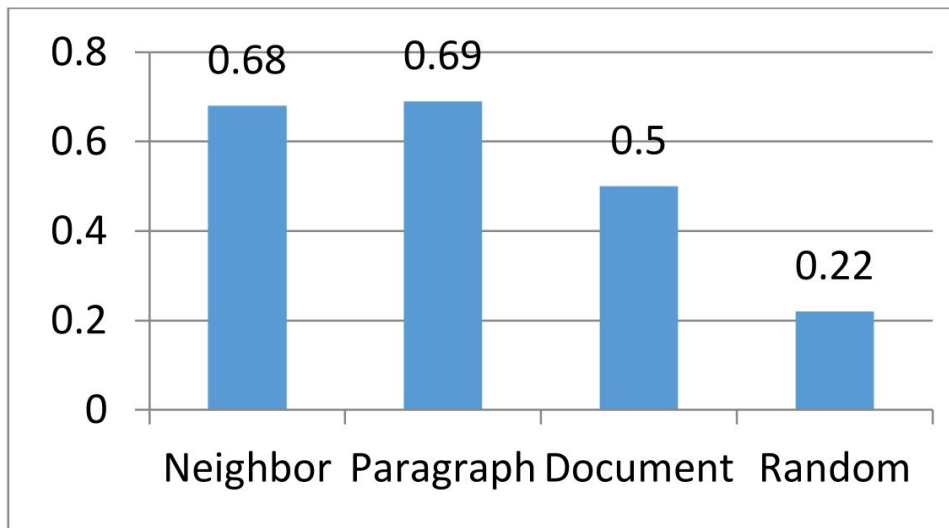
	#Sentence
No Label	180
One Label	1096
Two Labels	1081
Three Labels	346
Four or more labels	48
<b>ALL</b>	<b>2751</b>

# Korpusová data: závislost labelů



# Korpusová data: závislost na kontextu

- pravděpodobnost, že 2 věty vyjadřují stejnou emoci v různých prostředích



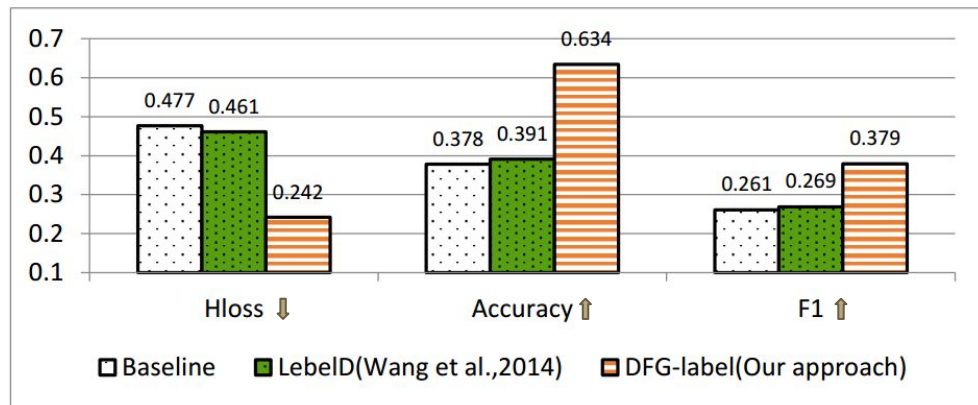
# Experiment s DFG

- 100 dokumentů celkem: 80 = trénovací, 20 = testovací
- Hodnotící metriky:
  - ↓ 1) **Hammingova vzdálenost** (Vyhodnocuje, kolikrát je dvojice chybně klasifikována s ohledem na předpokládaný výsledek a data získaná pozorováním.)
  - ↑ 2) **Přesnost** (Poskytuje průměrný stupeň podobnosti mezi předpokládanými výsledky a vyhodnocenými daty.)
  - ↑ 3) **F1-score** (Harmonický průměr přesnosti a pokrytí. Může být vypočítán z pravdivě pozitivních, pravdivě negativních, falešně pozitivních a falešně negativních výsledků na základě predikce a výsledku.)



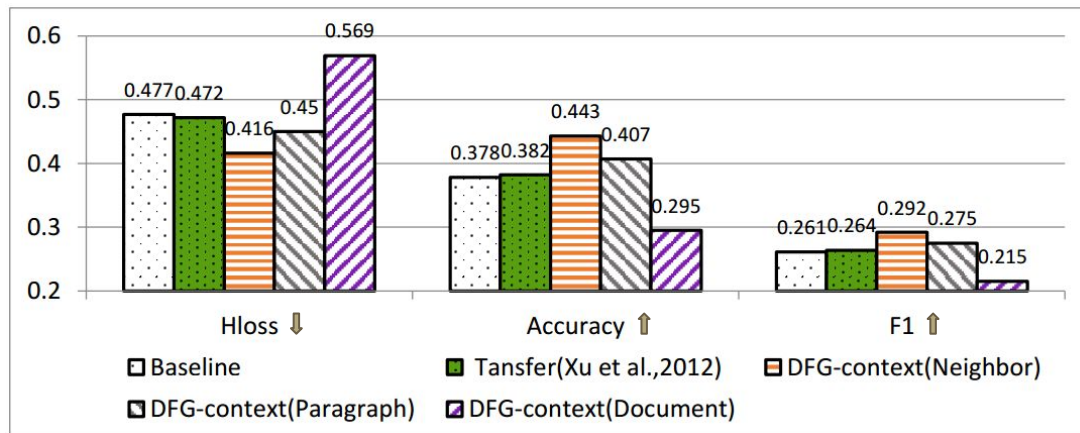
# Experiment s DFG: Závislost na labelu

- **baseline:** přístup, který uplatňuje maximální empirii, ale neohlíží se na závislost labelů ani kontextu
- **LabelD:** dosud nejlepší výzkum multi-label klasifikace, zahrnuje závislost labelů
- **DFG-label:** nový testovaný výzkum se závislostí labelů



# Experiment s DFG: Závislost na kontextu

- **Transfer:** nejlepší výzkum v oblasti informací o kontextu v klasifikaci emocí na úrovni věty
- **DFG-label (Neighbor/Paragraf/Document)**
  - klasifikace na úrovni dokumentu neúčinná × **neighbor**



# Experiment s DFG: Závislost labelů i na kontextu

- spojení obou závislostí zlepšuje výsledky
- kvalita není závislá na velikosti testovacích dat

	<i>Hloss</i>	<i>Accuracy</i>	<i>F1</i>
Baseline	0.447	0.378	0.261
DFG-label	0.254	0.621	0.372
DFG-context	0.416	0.443	0.292
DFG-both	<b>0.242</b>	<b>0.634</b>	<b>0.379</b>

# Zdroje

<http://www.aclweb.org/anthology/P/P15/P15-1101.pdf>

[https://en.wikipedia.org/wiki/Factor\\_graph](https://en.wikipedia.org/wiki/Factor_graph)

[https://cs.wikipedia.org/wiki/Bipartitní\\_graf](https://cs.wikipedia.org/wiki/Bipartitní_graf)

[https://en.wikipedia.org/wiki/F1\\_score](https://en.wikipedia.org/wiki/F1_score)

[https://cs.wikipedia.org/wiki/Hammingova\\_vzdálenost](https://cs.wikipedia.org/wiki/Hammingova_vzdálenost)

---

---

**Díky za pozornost**

*Daniela Ryšavá*

---

---