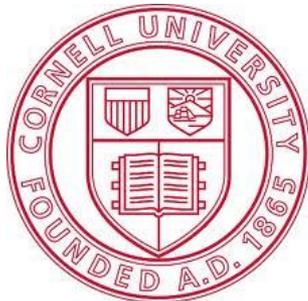


Thumbs up?

Sentiment Classification using Machine Learning Techniques

Авторы:

Bo Pang & Lillian Lee
Department of CS
Cornell University



Shivakumar
Vaithyanathan
IBM Almaden
Research Center



Докладчик:

А. Каминская
ИС ОПМИ
НИУ ВШЭ



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

Немного о статье: общее

Конференция EMNLP 2002 (Conference on Empirical Methods in Natural Language Processing)

«Thumbs up?
Sentiment Classification using Machine Learning
Techniques»

«Применение методов машинного обучения
для классификации мнений»



План

- Введение
- Релевантные работы
- Данные
- «Человеческая» классификация
- Методы машинного обучения
 - Наивный Байес
 - Максимальная энтропия
 - Метод опорных векторов
- Эксперименты, результаты
- Обсуждение, вопросы



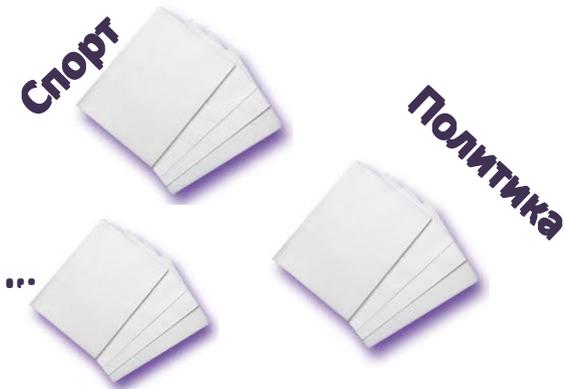
Введение в проблему



Огромное количество информации в онлайн документах;

Нужно «навести порядок» в этой информации, то есть классифицировать документы для пользователя.

Классификация по тематике



Классификация мнений



Предыдущие работы по теме

- Классификация по источнику (или типу источника) [Biber,1988]

The New York Times vs. The Daily News

- Классификация по жанру текста [Karlsgren and Cutting, 1994], [Kessler et al., 1997], [Finn et al., 2002]

review vs. abstract

- Классификация мнений [Hatzivassiloglou and McKeown, 1997; Turney and Littman, 2002]

Требуются заранее составленные списки определяющих слов

Данные



Рецензии на фильмы с явными оценками:

- **Позитивные**
- **Негативные**
- **Нейтральные**

Ограничение:

Не более 20 рецензий от автора (в одной категории)

Данные



Рецензии на фильмы с явными оценками:

- **Позитивные (1301 рецензия)**
- **Негативные (752 рецензии)**
- ~~Нейтральные~~

Ограничение:

Не более 20 рецензий от человека (в одной категории): итого 144 рецензента

«Человеческая» классификация

	Proposed word lists	Accuracy	Ties
Human 1	positive: <i>dazzling, brilliant, phenomenal, excellent, fantastic</i> negative: <i>suck, terrible, awful, unwatchable, hideous</i>	58%	75%
Human 2	positive: <i>gripping, mesmerizing, riveting, spectacular, cool, awesome, thrilling, badass, excellent, moving, exciting</i> negative: <i>bad, cliched, sucks, boring, stupid, slow</i>	64%	39%

Accuracy = % документов, классифицированных верно (точность) – чем выше, тем лучше

Ties = % документов, где оба мнения существовали равновероятно (связность) – чем ниже, тем лучше

	Proposed word lists	Accuracy	Ties
Human 3 + stats	positive: <i>love, wonderful, best, great, superb, still, beautiful</i> negative: <i>bad, worst, stupid, waste, boring, ?, !</i>	69%	16%

Список слов, основанный на «самоанализе» и небольшой статистике, показывает лучшие результаты



Самоанализ + статистика: вклад СТАТИСТИКИ

	Proposed word lists	Accuracy	Ties
Human 3 + stats	positive: <i>love, wonderful, best, great, superb, still, beautiful</i> negative: <i>bad, worst, stupid, waste, boring, ?</i>	69%	16%

«*Still, though, it was worth seeing*»

	Proposed word lists	Accuracy	Ties
Human 3 + stats	positive: <i>love, wonderful, best, great, superb, still, beautiful</i> negative: <i>bad, worst, stupid, waste, boring, ?</i>	69%	16%

«*What was the director thinking?*»

Методы машинного обучения

Цель: понять, эквивалентна ли классификация по тематике классификации мнений?

Или же нужны специальные методы?

Проверенные:

- Наивный Байесовский классификатор
- Метод максимальной энтропии
- Метод опорных векторов

Подготовка данных:

$\{f_1, \dots, f_m\}$ - множество признаков (напр., слов)

$\vec{d} := (n_1(d), n_2(d), \dots, n_m(d))$. - вектор \forall документа

Методы МО: Наивный Байес



Основное предположение: признаки документа независимы!

Суть: определить для данного документа d подходящий класс c^* :

$$c^* = \operatorname{argmax}_c P(c|d)$$

$$P(c | d) = \frac{P(c)P(d | c)}{P(d)}$$

$$P_{\text{NB}}(c | d) := \frac{P(c) \left(\prod_{i=1}^m P(f_i | c)^{n_i(d)} \right)}{P(d)}$$

Методы МО: максимальная энтропия

ME иногда превосходит NB;

Не нужны ограничения на признаки документов!

Суть: определить для данного документа d подходящий класс c^* :

$$c^* = \operatorname{argmax}_c P(c|d)$$

$$P_{\text{ME}}(c | d) := \frac{1}{Z(d)} \exp \left(\sum_i \lambda_{i,c} F_{i,c}(d, c) \right)$$

нормализация

$$Z(d) = \sum_c \exp \left(\sum_i \lambda_i f_i(d, c) \right)$$

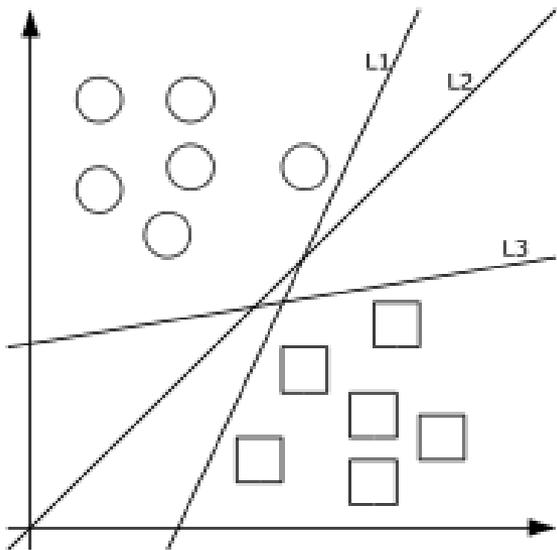
веса (чем выше,
тем лучше)

$$F_{i,c}(d, c') := \begin{cases} 1, & n_i(d) > 0 \text{ and } c' = c \\ 0 & \text{otherwise} \end{cases}$$

Методы МО: метод опорных векторов

Support Vector Machines (SVM) обычно превосходит NB в традиционной классификации документов.

Суть: найти разделяющую гиперплоскость

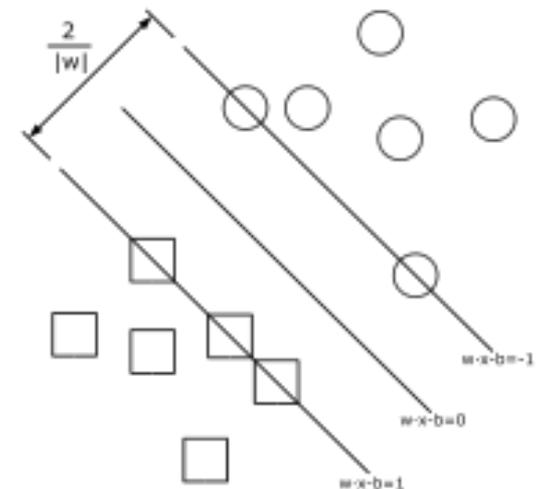


$$c_j \in \{1, -1\}$$

$$\vec{w} := \sum_j \alpha_j c_j \vec{d}_j, \quad \alpha_j \geq 0$$



опорные вектора (если $\alpha_j > 0$)



Эксперименты

- 700 позитивных + 700 негативных рецензий (корпус из 1400 текстов)
- Деление выборки на 3 «равномерные» части
- «unigram» - слово содержится в корпусе минимум 4 раза: 16165
- «bigram» - 2 слова содержатся в корпусе минимум 7 раз: 16165
- «POS» - пометки частей речи («*I love this story*» vs. «*This is a love story*»)

Результаты

	Features	# of features	frequency or presence?	NB	ME	SVM
(1)	unigrams	16165	freq.	78.7	N/A	72.8
(2)	unigrams	”	pres.	81.0	80.4	82.9
(3)	unigrams+bigrams	32330	pres.	80.6	80.8	82.7
(4)	bigrams	16165	pres.	77.3	77.4	77.1
(5)	unigrams+POS	16695	pres.	81.5	80.4	81.9
(6)	adjectives	2633	pres.	77.0	77.7	75.1
(7)	top 2633 unigrams	2633	pres.	80.3	81.0	81.4
(8)	unigrams+position	22430	pres.	81.0	80.1	81.6



Результаты



	Features	# of features	frequency or presence?	NB	ME	SVM
(1)	unigrams	16165	freq.	78.7	N/A	72.8
(2)	unigrams	”	pres.	81.0	80.4	82.9
(3)	unigrams+bigrams	32330	pres.	80.6	80.8	82.7
(4)	bigrams	16165	pres.	77.3	77.4	77.1
(5)	unigrams+POS	16695	pres.	81.5	80.4	81.9
(6)	adjectives	2633	pres.	77.0	77.7	75.1
(7)	top 2633 unigrams	2633	pres.	80.3	81.0	81.4
(8)	unigrams+position	22430	pres.	81.0	80.1	81.6

Обсуждение

МО для классификации мнений гораздо лучше в смысле точности, чем «ручная» классификация!

Исключения:

*“This film should be **brilliant**. It sounds like a **great** plot, the actors are **first grade**, and the supporting cast is **good** as well, and Stallone is attempting to deliver a **good** performance. **However, it can't hold up**”*

*“I **hate** the Spice Girls. ...[3 things the author hates about them]... Why I saw this movie is a really, really, really long story, but I did, and one would think I'd **despise** every minute of it. But... Okay, I'm really ashamed of it, but I **enjoyed** it. I mean, I admit it's a really **awful** movie ...**the ninth floor of hell**...The plot is such a **mess** that it's **terrible**. But I **loved** it.”*

Спасибо!
Вопросы?

