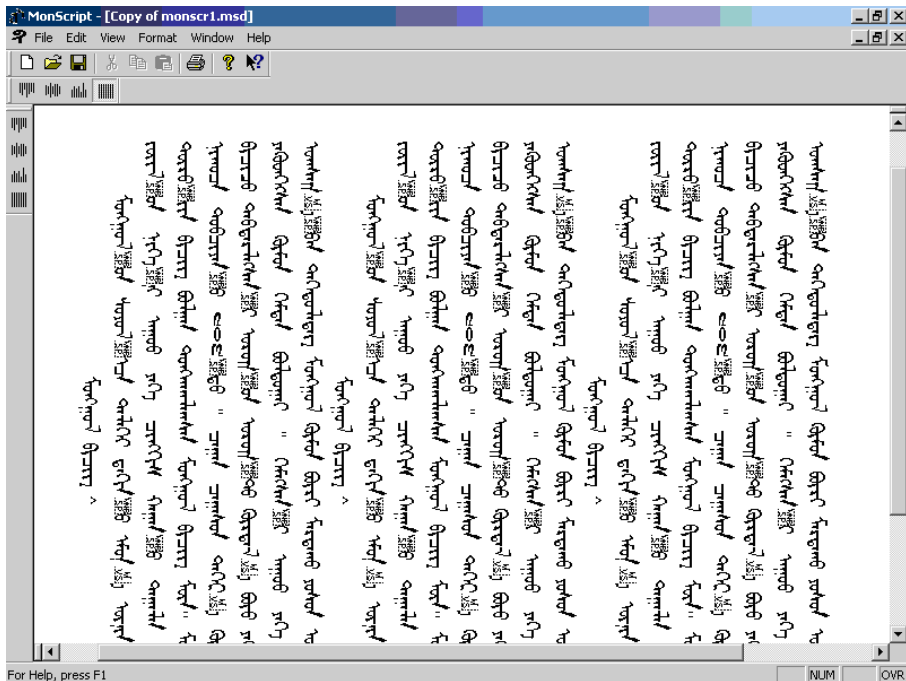


Хандлагын шинжилгээ

Хандлагын
шинжилгээ гэж юу
вэ?



Киноны эерэг эсвэл сөрөг шүүмж?



- unbelievably disappointing



- Full of zany characters and richly applied satire, and some great plot twists



- this is the greatest screwball comedy ever filmed



- It was pathetic. The worst part about it was the boxing scenes.

Google дээрх бүтээгдэхүүн хайлт



HP Officejet 6500A Plus e-All-in-One Color Ink-jet - Fax / copier / printer / scanner

\$89 online, \$100 nearby ★★★★★ 377 reviews

September 2010 - Printer - HP - Inkjet - Office - Copier - Color - Scanner - Fax - 250 shi

Reviews

Summary - Based on 377 reviews



What people are saying

ease of use	<div><div></div><div></div></div>	"This was very easy to setup to four computers."
value	<div><div></div><div></div></div>	"Appreciate good quality at a fair price."
setup	<div><div></div><div></div></div>	"Overall pretty easy setup."
customer service	<div><div></div><div></div></div>	"I DO like honest tech support people."
size	<div><div></div><div></div></div>	"Pretty Paper weight."
mode	<div><div></div><div></div></div>	"Photos were fair on the high quality mode."
colors	<div><div></div><div></div></div>	"Full color prints came out with great quality."

Bing Shopping

HP Officejet 6500A E710N Multifunction Printer

[Product summary](#) [Find best price](#) **Customer reviews** [Specifications](#) [Related items](#)



\$121.53 - \$242.39 (14 stores)

☐ Compare

Average rating ★★★★★ (144)



Most mentioned

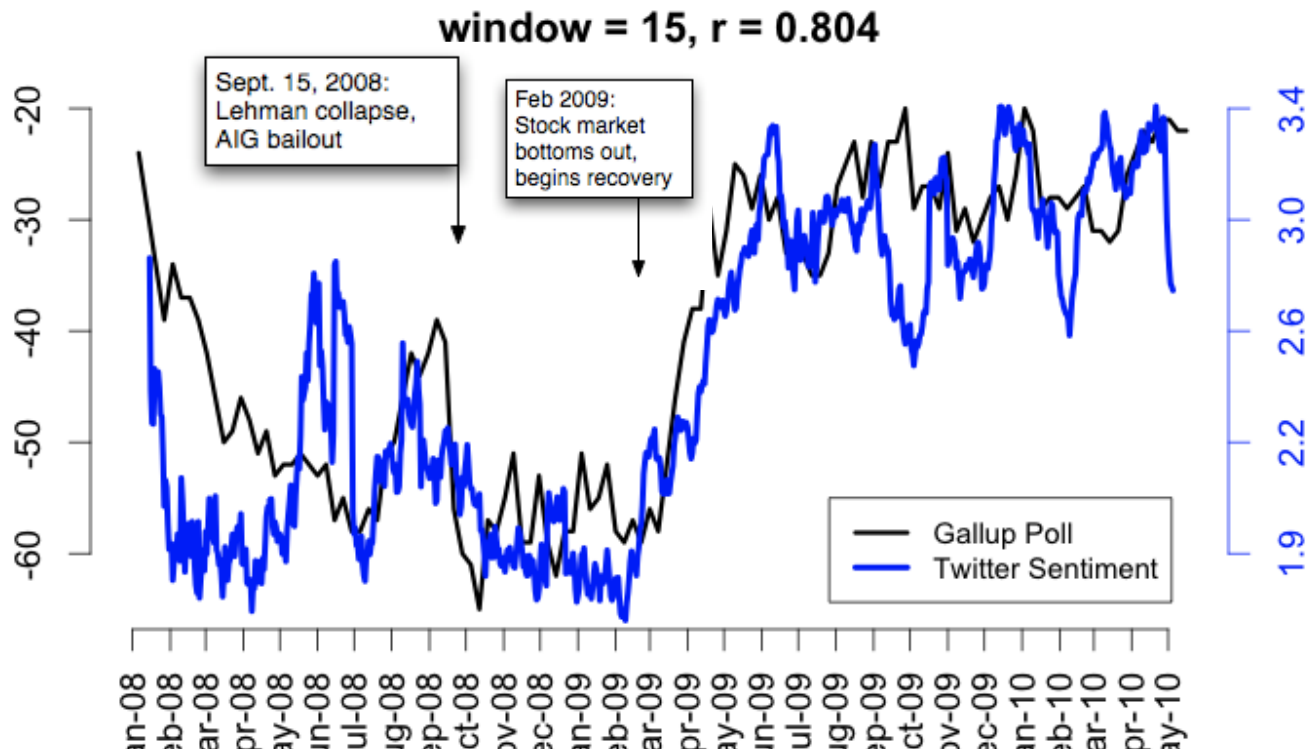


Show reviews by source

Best Buy (140)
CNET (5)
Amazon.com (3)

Twitter –ийн хандлагын эсрэг Gallup компаний сэтгэл ханамжийн санал асуулга

Brendan O'Connor, Ramnath Balasubramanian, Bryan R. Routledge, and Noah A. Smith. 2010.
From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In ICWSM-2010



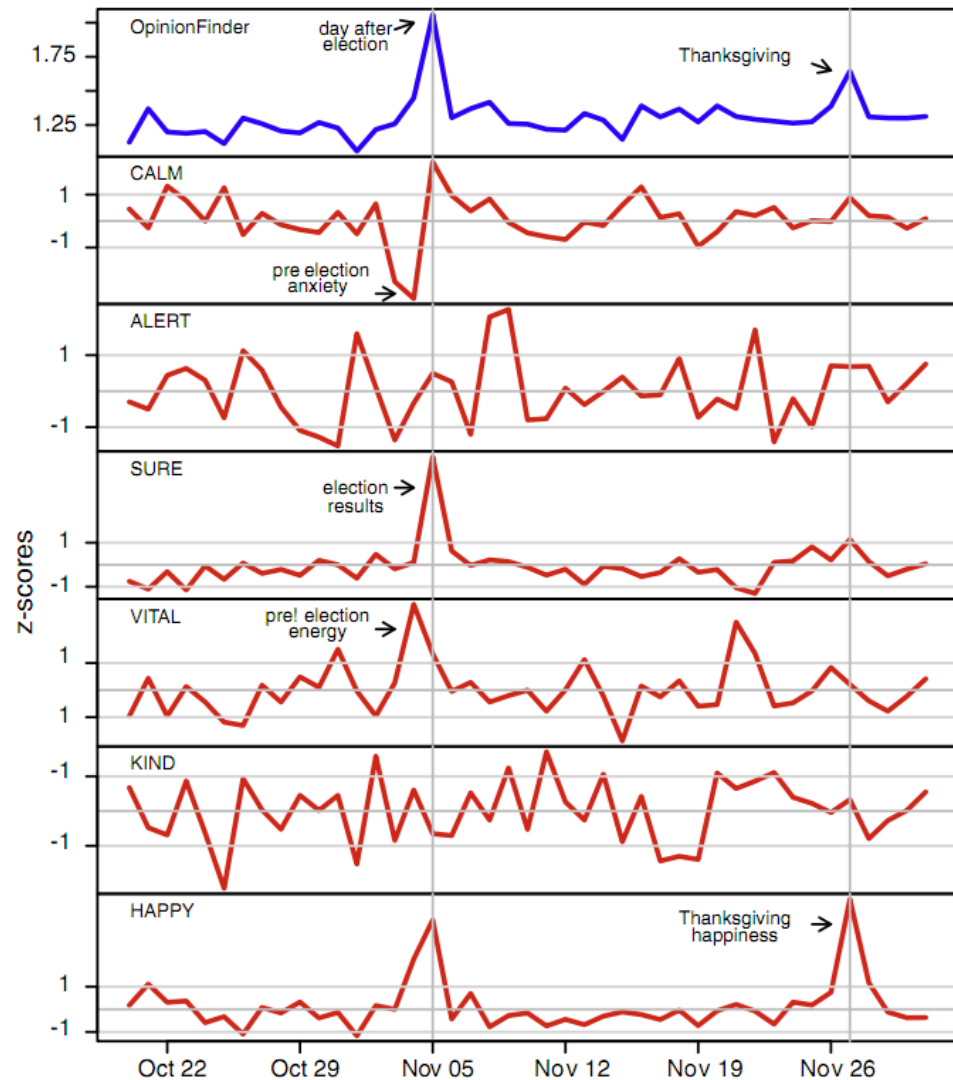
Twitter-ийн хандлага

Johan Bollen, Huina Mao, Xiaojun Zeng. 2011.

Twitter mood predicts the stock market,

Journal of Computational Science 2:1, 1-8.

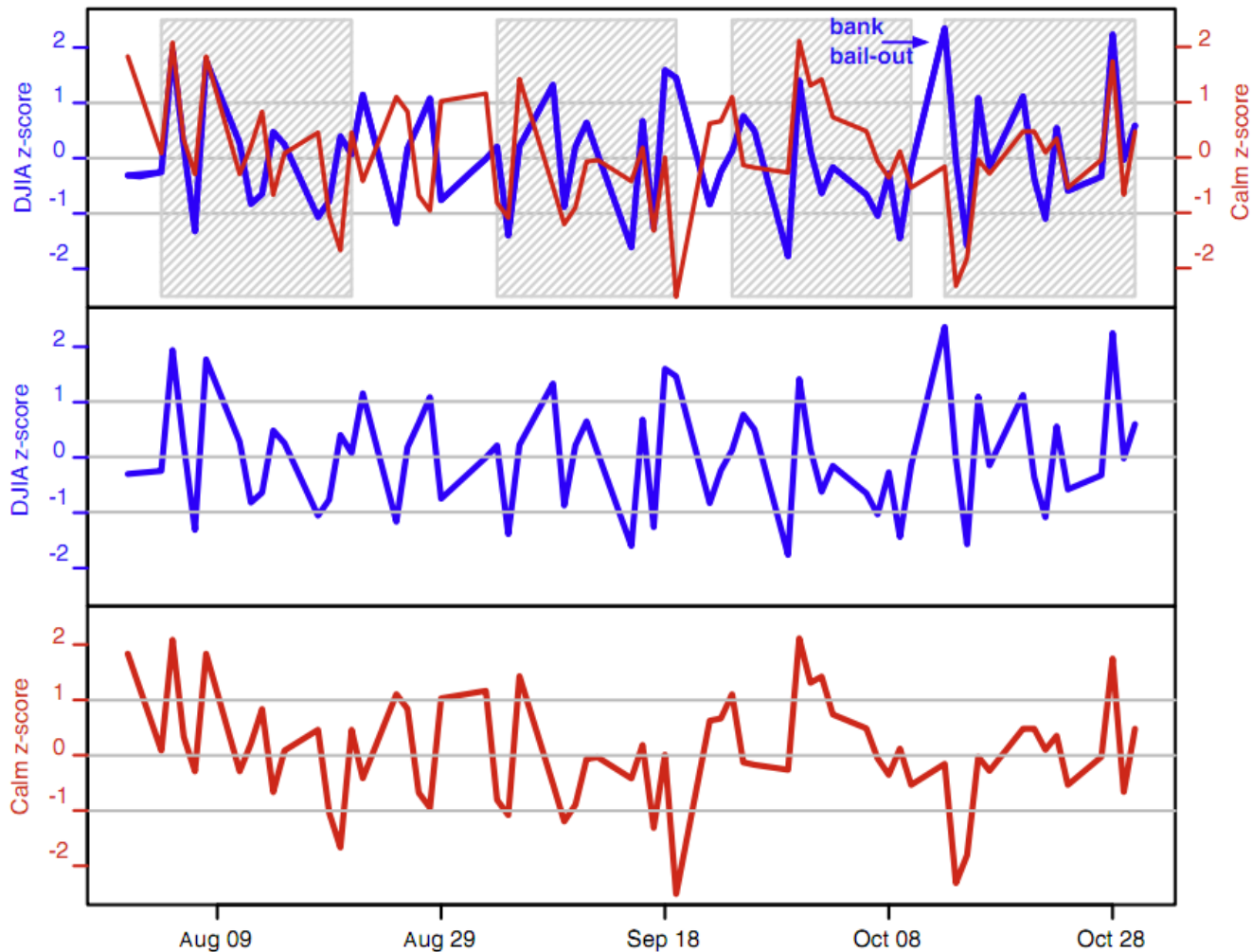
10.1016/j.jocs.2010.12.007.



Bollen et al. (2011)

- CALM 3 хоногийн дараа DJIA-г урьдчилан таамагдаж байна
- Ядаж нэг хөрөнгө оруулалтын сан энэ алгоритмыг ашигладаг

CALM Dow Jones



Twitter дээрх зорилтот хандлага

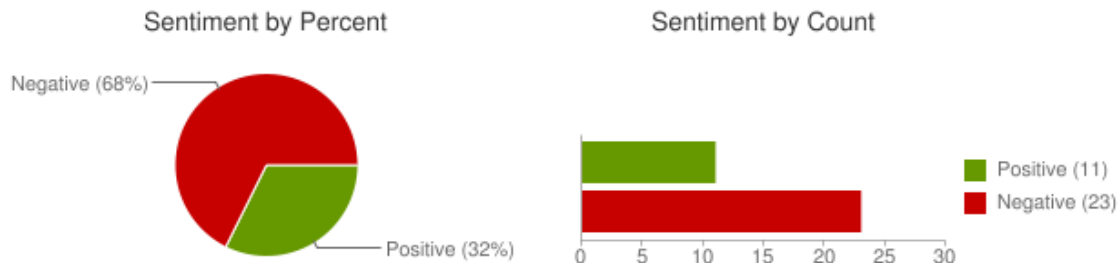
- Twitter -ийн хандлагын апп

- Alec Go, Richa Bhayani, Lei Huang. 2009. Twitter Sentiment Classification using Distant Supervision

Type in a word and we'll highlight the good and the bad

[Save this search](#)

Sentiment analysis for "united airlines"



jljacobson: OMG... Could @United airlines have worse customer service? W8g now 15 minutes
Posted 2 hours ago

12345clumsy6789: I hate United Airlines Ceiling!!! Fugn impossible to get my conduit in this d...
Posted 2 hours ago

EMLandPRGbelgiu: EML/PRG fly with Q8 united airlines and 24seven to an exotic destination
Posted 2 hours ago

CountAdam: FANTASTIC customer service from United Airlines at XNA today. Is tweet more,
Posted 4 hours ago

Хандлагын шинжилгээнд өөр олон нэршил байдаг

- Үзэл бодол олборлох,
- Үзэл бодол гаргах,
- Хандлага олборлох,
- Субъектив шинжилгээ.

Яагаад хандлагын шинжилгээ гэж?

- *Кино*: энэ шүүмж эерэг үү эсвэл сөрөг үү?
- *Бүтээгдэхүүн*: энэ шинэ iPhone-ний талаар хүмүүс ямар бодолтой байна?
- *Олон нийтийн хандлага*: хэрэглэгчийн итгэл ямар байна? Итгэлгүй байдал нэмэгдэж үү?
- *Улс төр*: хүмүүс нэр дэмшигчийн талаар эсвэл асуудлын талаар ямар бодолтой байна?
- *Prediction*: сонгуулийн үр дүн эсвэл зах зээлийн хандлагыг

Сэтгэл хөдлөлийн төлөвийн хэв шинжүүд

- **Эмоци:** үйл явдлын товч үнэлгээ
 - *ууртай, гунигтай, баяр хөөртэй, айдас, ичгүүртэй, бардам, баярласан*
- **Сэтгэл санаа:** Субъектив мэдрэмжийн өөрчлөлтөөс үүдэлтэй бага эрчимтэй удаан хугацааны өөрчлөлт
 - *хөгжилтэй, уйтгартай, уур уцаартай, жагсаалгүй, сэтгэлээр унасан, ууртай*
- **Хувь хүн хоорондын хандлага:** тодорхой харилцаанд байгаа өөр хүнтэй харьцах хандлага
 - *найрсаг, сээтэгнэх, алс холын, хүйтэн, дулаан, дэмждэг, үл тоомсорлодог*
- **Хандлага:** объект эсвэл хүн рүү хандах хандлага, тогтсон сэтгэл хөдлөл бүхий үзэл бодол
 - *таалагдах, хайрлах, үзэн ядах, үнэлэх, хүсэх*
- **Хувийн зан төлөв:** тогтвортой зан чанар, зан үйлийн ердийн хандлага
 - *бухимдалтай, түгшүүртэй, болгоомжгүй, ааш муутай, дайсагнасан, атаархдаг*

Сэтгэл хөдлөлийн төлөвийн хэв шинжүүд

- Эмоци: үйл явдлын товч үнэлгээ
 - ууртай, гунигтай, баяр хөөртэй, айдас, ичгүүртэй, бардам, баярласан
- Сэтгэл санаа: Субъектив мэдрэмжийн өөрчлөлтөөс үүдэлтэй бага эрчимтэй удаан хугацааны өөрчлөлт
 - хөгжилтэй, уйтгартай, уур уцаартай, жагсаалгүй, сэтгэлээр унасан, ууртай
- Хувь хүн хоорондын хандлага: тодорхой харилцаанд байгаа өөр хүнтэй харьцах хандлага
 - найрсаг, сээтэгнэх, алс холын, хүйтэн, дулаан, дэмждэг, үл тоомсорлодог
- **Хандлага:** объект эсвэл хүн рүү хандах хандлага, тогтсон сэтгэл хөдлөл бүхий үзэл бодол
 - таалагдах, хайрлах, үзэн ядах, үнэлэх, хүсэх
- Хувийн зан төлөв: тогтвортой зан чанар, зан үйлийн ердийн хандлага
 - бүхимдалтай, түгшүүртэй, болгоомжгүй, ааш муутай, дайсагнасан, атаархдаг

Хандлагын шинжилгээ

- Хандлагын шинжилгээ нь **хандлага** тодорхойлдог.
“объект эсвэл хүн рүү хандах хандлага, тогтсон сэтгэл хөдлөл бүхий үзэл бодол”
 1. Хандлагын **эзэн (эх сурвалж)**
 2. Хандлагын **зорилго (үзэл бодол)**
 3. Хандлагын **төрөл**
 - Бүлэг төрөлөөс
 - *дуртай, хайр, үзэн ядах, үнэ цэнэ, хүсэл, гэх мэт.*
 - Эсвэл (илүү нийтлэг) энгийн жигнэсэн **туйлшрал**:
 - *ээрэг, сөрөг, төвийг сахьсан, хүчтэй* хамт олон
 4. **Текст** нь хандлагыг агуулсан байна
 - Өгүүлбэр эсвэл бүтэн документ

Хандлагын шинжилгээ

- Энгийн даалгавар:
 - Текстийн хандлага эерэг үү сөрөг үү?
- Илүү түвэгтэй:
 - Хандлагын зэрэг 1 -ээс 5
- Дэвшилтэт:
 - Зорилго, эх сурвалжийг таних, эсвэл хандлагын нүсэр төрлүүд

Хандлагын шинжилгээ

- Энгийн даалгавар:
 - Текстийн хандлага эерэг үү сөрөг үү?
- Илүү түвэгтэй:
 - Хандлагын зэрэг 1 -ээс 5
- Дэвшилтэт:
 - Зорилго, эх сурвалжийг таних, эсвэл хандлагын нүсэр төрлүүд

Кино шүүмж дэх хандлагын ангилал

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP-2002, 79—86.

Bo Pang and Lillian Lee. 2004. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. ACL, 271-278

- Туйлшралын таних:
 - IMDB (Internet movie database) кино шүүмж эерэг үү эсвэл сөрөг үү?
- Өгөгдөл: *Polarity Data 2.0*:
 - <http://www.cs.cornell.edu/people/pabo/movie-review-data>

Pang болон Lee өгөгдлийн сан дах IMDB өгөгдөл



when _star wars_ came out some twenty years ago , the image of traveling throughout the stars has become a commonplace image . [...]

when han solo goes light speed , the stars change to bright lines , going towards the viewer in lines that converge at an invisible point .

cool .

october sky offers a much simpler image—that of a single white dot , traveling horizontally across the night sky . [. . .]



“ snake eyes ” is the most aggravating kind of movie : the kind that shows so much potential then becomes unbelievably disappointing .

it's not just because this is a brian depalma film , and since he's a great director and one who's films are always greeted with at least some fanfare .

and it's not even because this was a film starring nicolas cage and since he gives a brauvara performance , this film is hardly worth his talents .

Үндсэн алгоритм (Pang болон Lee нарын авсан)

- Токенчлол
- Онцлог шинж олборлолт
- Ялгаатай ангилагчид ашигласан ангилал
 - Найв Бейс
 - Максимум энтропи
 - Тулгуур векторын арга

Хандлага дээр токенчлох асуудал

- HTML ба XML тэмдэглэгээтэй тулгардаг
- Twitter –ийн тэмдэглэгээ (нэр, хэйш шошго) Potts –ийн эмоциуд
- Том үсгээр хадгалах
- Утасны дугаар, огноо
- Эмоци
- Хэрэгтэй код:
 - Christopher Potts хандлагын токенчлогч
 - Brendan O'Connor twitter-ийн токенчлогч

```
[<>]? # optional hat/brow
[:;=8] # eyes
[\\-o\\*\\']? # optional nose
[\\)\\]\\(\\[dDpP/\\:~\\}\\{\\@\\|\\|\\] # mouth
| ### reverse orientation
[\\)\\]\\(\\[dDpP/\\:~\\}\\{\\@\\|\\|\\] # mouth
[\\-o\\*\\']? # optional nose
[:;=8] # eyes
[<>]? # optional hat/brow
```

Хандлагын ангилалд онцлог шинжүүдийг олборлох нь

- Үгүйсгэлийг хэрхэн шийдэх вэ?
 - I **didn't** like this movie
эсрэгээр
 - I really like this movie
- Аль үгсийг ашиглах вэ?
 - Зөвхөн тэмдэг нэр
 - Бүх үг
 - Бүх үг байхад илүү сайн ажилладаг

Үгүйсгэл

Das, Sanjiv and Mike Chen. 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. In Proceedings of the Asia Pacific Finance Association Annual Conference (APFA).
Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP-2002, 79—86.

Үгүйсгэл болон дараагийн цэг таслалын хоорондын үг бүрт NOT_ -ийг
НЭМ:

didn't like this movie , but I



didn't NOT_like NOT_this NOT_movie but I

Эргэн санахад: Найв Бейс

$$c_{NB} = \underset{c_j \in C}{\operatorname{argmax}} P(c_j) \prod_{i \in \text{positions}} P(w_i | c_j)$$

$$\hat{P}(w | c) = \frac{\text{count}(w, c) + 1}{\text{count}(c) + |V|}$$

2-тын (Булийн шинж чанартай) олон гишүүнт Найв Бейс

- Санаа нь:
 - Хандлагад (болон бусад текст ангиллын магадлалд)
 - Үг тохиолдох нь үгийн давтамжаас илүү чухал байж болно
 - *fantastic* гэсэн үг тохиолдох нь бидэнд их мэдээлэл өгдөг
 - Гэтэл тэр 5 удаа тохиолдох нь тийм чухал биш.
 - Булийн олон гишүүнт Найв Бейс
 - Документ бүрт бүх үгийг 1-ээр хадгална

Булийн олон гишүүнт Найв Бейс: Сургалт

- Сургалтын корпусаас, *Үгийн сан* олборлоно

- $P(c_j)$ томъёог тооцоол

- For each c_j in C do

$docs_j \leftarrow$ all docs with class = c_j

$$P(c_j) \propto \frac{|docs_j|}{|\text{total \# documents}|}$$

- $P(w_k | c_j)$ томъёог тооцоол

- Remove duplicates in containing all $docs_j$

- For each word type w in $docs_j$

- Retain only a single instance of w

$n_k \leftarrow$ # of occurrences of w_k in $Text_j$

$$P(w_k | c_j) \propto \frac{n_k + a}{n + a | \text{Vocabulary} |}$$

Тестийн d документ дэх Булийн олон гишүүнт Найв Бейс

- Эхлээд d –ээс бүх давхардсан үгсийг хасна
- Дараа нь НБ дараах томъёогоор тооцоолно:

$$c_{NB} = \underset{c_j \in C}{\operatorname{argmax}} P(c_j) \prod_{i \in \text{positions}} P(w_i | c_j)$$

Энгийн vs. Булийн олон гишүүнт НБ

Энгийн	Док	Үгс	Анги
Сургалт	1	Chinese Beijing Chinese	c
	2	Chinese Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Тест	5	Chinese Chinese Chinese Tokyo Japan	?

Булийн	Док	Үгс	Анги
Сургалт	1	Chinese Beijing	c
	2	Chinese Shanghai	c
	3	Chinese Macao	c
	4	Tokyo Japan Chinese	j
Тест	5	Chinese Tokyo Japan	?

2-тын (Булийн шинж чанартай) олон гишүүнт Найв Бейс

B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. EMNLP-2002, 79—86.

V. Metsis, I. Androutsopoulos, G. Paliouras. 2006. Spam Filtering with Naive Bayes – Which Naive Bayes? CEAS 2006 - Third Conference on Email and Anti-Spam.

K.-M. Schneider. 2004. On word frequency information and negative evidence in Naive Bayes text classification. ICANLP, 474-485.

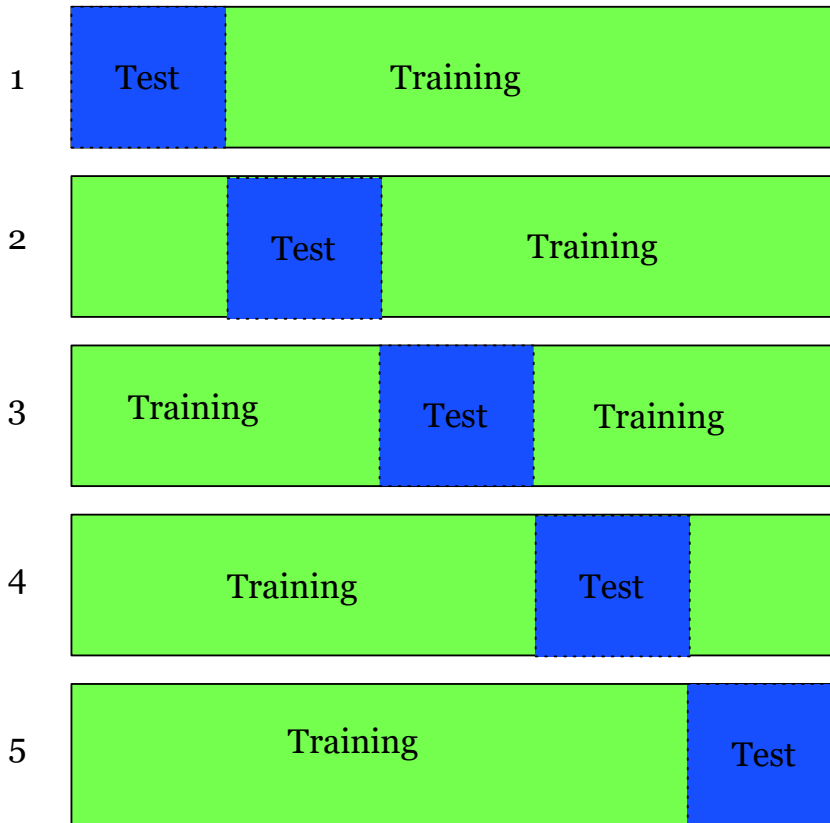
JD Rennie, L Shih, J Teevan. 2003. Tackling the poor assumptions of naive bayes text classifiers. ICML 2003

- 2-тын нь үгийн давтамжийн тоотойгоос илүү ажиллах мэт харагддаг
 - Энэ нь Олон хувилбарт НБ шиг ижил **биш**
 - ОХНБ нь хадлага эсвэл бусад даалгавруудад сайн ажилладаггүй
- Бусад боломж: $\log(\text{freq}(w))$

Хөндлөнгийн үнэлгээ

- Өгөгдлийг 10 нугалаанд хуваана
 - (Нугалаа бүр дотор хагас нь эерэг, хагас нь сөрөг?)
- Нугалаа бүрийн хувьд
 - түр зуурын тестийн олонлогийг сонгоно
 - 9 нугалаа дээр сургаж, тестийн олонлог дээр гүйцэтгэлийг тооцоолно
- 10 ажиллагааны дундаж гүйцэтгэлийг хэвлэнэ

Iteration



Ангилалын бусад асуудлууд

- Хамгийн их энтропи ба Тулгуур вектор машин нь Найв Бейсийг бодвол илүү сайн байх хандлагатай байдаг.

Асуудлууд:

Юу шүүмжийг ангилахад хэцүү болгодог вэ?

- Ухаан шаардсан бичиглэл:
 - Үнэртэй уснаас хийсэн сүрчигний шүүмж:
 - “If you are reading this because it is your darling fragrance, please wear it at home exclusively, and tape the windows shut.”
 - Кэтрин Хепберн дээр Дороти Паркер:
 - “She runs the gamut of emotions from A to B”

Хийсвэр хүлээлт ба онцгүй үр дүн

- “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.”
- Well as usual Keanu Reeves is nothing special, but surprisingly, the very talented Laurence Fishbourne is not so good either, I was surprised.