Lecture 10 - Case

Sondre Elstad

2022-06-25

Introduction

In this final lecture we will consider the article "The Value of News" (written by Vegard H. Larsen and Leif Anders Thorsrud in 2015). Or, rather, we will consider a part of the article. This is certainly not an economics class - although I would like to take this opportunity to wholeheartedly recommend this article. In a way, this is related to the political science-article we talked about a couple of lectures ago. It is related in the sense that this article also considers textual data. In particular, it employs a so-called Latent Dirichlet Allocation model to facilitate the process of crafting a news index.

News and expectations

Changes in expectations may affect the economy. If bad news comes, this might cause households to reduce their consumption and increase their savings. It might cause firms to postpone their investments. Goods news, on the other hand, can have quite the contrary effects. We currently know too little about this to establish clear causal links. It is, however, consistent with a lot of ideas in economics. Decision-making under uncertainty is certainly something we would like to learn more about. We can think of it as a signal extraction exercise. When agents manage to extract positive signals correctly and act on them, the economy will boom. When the news, instead of being substantial, proves to be noise, this might have a short-term effect on the economy. But the effect will fade as agents revise their expectations. True news should thus have predictive power for the future developments of the economy. The trouble is that what is 'new' (in the sense that it is unexpected) news and not is not something we can observe.

Latent Dirichlet Allocation (LDA)

Larsen & Thorsrud (2015) extracts articles from Dagens Næringsliv (DN). DN is a Norwegian business newspaper. The articles are available at Retriever's "Atekst" database. It gives access to DN articles published from May 2 1988 to December 29 2014. We are talking a total of 459 745 articles, more than one billion words and more than a million unique tokens. It's a sample of over 9000 days. They do all of the regular preprocessing: stemming, lemmatization, removal of stop-words etc. They also measure the tf-idf (term frequency - inverse document frequency). They keep around 250 000 of the stems with the highest tf-idf score and use this as their final corpus.

The LDA is an unsupervised learning algorithm. It clusters words into topics. A topic is a distribution over words. At the same time, it classifies articles as mixtures of topics. Words convey information. However, it is the information we are interested in. We are looking for something that is latent in the words. Finally, the term "Dirichlet" is used because the topic mixture is drawn from a conjugate Dirichlet prior. Hence the name Latent Dirichlet Allocation.

There is no correct way to assign words to topics, but Larsen & Thorsrud (2015) provide the following reasonable list:

Topic	Label	# of articles	First words
Topic 0	Anglo-Saxon	4457	the, new, of, york, doll, and, in, london, world, street, is, you, on, english, england, wall
Topic 1	Leadership	4357	position, forests, chairman, president, ceo, dismissal, executive, candidate, elected
Topic 2	Unknown	5717	smile, night, man, wall, house, door, gate, clock, home, minute, no, night, black
Topic 3	Knowledge	707	know, things, think, answer, never, good, feel, always, really, need, tell, pretty, just, feel, try
Topic 4	Context	710	degree, power, unequal, change, influence, context, difference, high, impact, significantly
Topic 5	Public safety	7497	police, finance guards, sight, illegal, investigation, indicted, prison, corruption, report
Topic 6	Government policy	3948	suggestions, parliamentary, department, selection, treasury, minister, change, foss, budget
Topic 7	Olympics	2301	olymics, participate, visit, invite, lillehammer, interest, business, gold, arranging, walk, story
Topic 8	Cooperation	1435	group, cooperation, establish, trap, tandberg, strategy, data, ulltveit, develop, abb, alliances
Topic 9	Manufacturing	6850	product, production, produce, factory, manufacturer, brand, bet, competition, marketing
Topic 10	Support	1638	support, establish, organize, funding, culture, advice, help, freely, purposes, create, shape
Topic 11	Sweden	4894	swedish, sweden, nordic, north, stockholm, finland, finnish, ericsson, denmark, ab, island
Topic 12	Stock market	9519	exchange, fell, quotes, steps, investor, stockmarket, index, points, upswing, decreasing
Topic 13	Automobiles	6991	car, model, engine, drive, ford, volvo, toyota, mercedes, bmw, class, saab, brand
Topic 14	Funding	4916	loans, interest, equity, guarantee, funding, finance, financial, bond, risk, financial crisis
Topic 15	Employment benefits	5968	public, private, scheme, sector, pension, payday, measures, labor, working, service
Topic 16	Art	7169	picture, art, exhibition, gallery, artists, museum, munch, painting, auction, design
Topic 17	Sport	4815	games, game, club, soccer, sponsoring, sports, world cup, cosmopolitan, skiing, jahr, win

Topic	Label	# of articles	First words
Topic 18	Europe	6082	german, germany, european, french, euro, france, spain, italy, spanish, berlin, italian
Topic 19	IT/Technology	8872	internet, online, technology, pc, microsoft, service, system, electronic, apple, user, machine
Topic 20	Conflict	5978	war, iraq, military, attack, forces, al, conflict, defense, iran, israel, nato, soldier, un, vest
Topic 21	Success	411	top, list, space, good, happy, road, number, eight, seven, loud, close, joy, promise, right
Topic 22	Communication	7701	telenor, mail, mobile, customer, netcom, hermans, telia, online, vimpelcom, telecom
Topic 23	Brokerage firms	4442	customer, brokerage, trading, bonus, trade, securities, brokerage, acta, industry
Topic 24	Reasoning	774	should, therefore, quite, moreover, faith, sure, namely, right, of course, interesting, hardly
Topic 25	Family	4597	woman, children, men, family, young, father, man, home, mother, age, parents, age, son
Topic 26	Food	4913	wine, food, restaurant, taste, salt, nok, pepper, drinks, fruit, fresh, bottle, menu, server
Topic 27	Investigation	1103	report, investigate, analysis, conclusions, assessment, conducted, conclude, answers, base
Topic 28	Shipping	12441	ships, shipping, dollar, wilhelms, fleet, proud, frontliners, berges, tank, rat, skaug, ugland
Topic 29	Criticism	2886	criticism, express, asserting, article, claim, incorrectly, press, pr, react, should, respond
Topic 30	LLC	4120	llc, group, family, dividend, asset, holding, equity, subsidiary, ownership, shareholder
Topic 31	East Asia	8142	china, japan, chinese, asia, japanese, indians, dollar, government, kong, brazil, korea, south
Topic 32	Aviation	8951	sas, fly, travel, airline, english, braathens, airport, passenger, gardemoen, color, air, traffic
Topic 33	IT/Startup	5112	it, group, acquisitions, partner, establish, business, entrepreneur, steen, industry, office
Topic 34	UK/US presidents	5428	british, london, president, uk, election, pound, bush, obama, political, clinton, conservative
Topic 35	Monetary policy	11863	interest, central bank, inflation, point, governor, percentage points, fell, steps, economy
Topic 36	Industry	4001	industry, industries, workplace, business, create, small, competition, help, better, develop
Topic 37	Rig	8420	issue, rig, dollar, offshore, collier, shareholder, drilling, retrieve, seadrill, sundal, pareto
Topic 38	Life	4551	human, history, words, live, feel, kind, shape, death, man, old, him, never, express, modern

Topic	Label	# of articles	First words
Topic 39	Newspapers	10603	newspaper, media, press, schibsted, Dagbladet, journalist, vg, eve mail, editor
Topic 40	Negotiations	684	solution, negotiation, agreement, parties, confirm, offers, conversation, process, negotiate
Topic 41	EU	8997	eu, ef, eea, commission, membership, no, brussel, eft, farmers, negotiations, agriculture
Topic 42	TV	10852	television, nrk, channel, advertising, radio, digital, media, agency, program, commercial
Topic 43	Financial supervision	4290	letter, information, financial supervision, enlightenment, auditors, control, accounting
Topic 44	Oil production	10415	statoil, hydro, oil, field, gas, oil company, shelf, stavanger, platform, shell, findings
Topic 45	Charity	3515	south, organization, africa, church, poor, help, congo, red, aid, rich, big, un, trade council
Topic 46	Justice	6206	lawyer, judge, appeals, claims, supreme court, claim, lawsuit, district court, strife, legal
Topic 47	Literature	7430	book, read, books, reading, writing, history, novel, writer, no, name, him, acted, author
Topic 48	Calender	725	week, previous, january, march, monday, friday, october, december, november, february
Topic 49	Aker	7591	aker, kværner, røkke, finance, option, tdn, rgi, shareholder, hafslund, enlighten, cruise
Topic 50	Projects	2353	project, cost, investment, cover, construction, operation, expansion, budget, annual
Topic 51	Nature	4314	water, meter, city, boat, mountains, ocean, outside, accident, weather, human, earth
Topic 52	Denmark	2017	danish, foreign, denmark, norwegians abroad, immigration, copenhagen, outdoors
Topic 53	Fishery	7457	fish, salmon, tons, seafood, food, marine, fishing, pan, fjord, norway, boat, plant, kilo
Topic 54	Retail	9795	shop, hotel, brand, trondheim, hotel, rema, reitan, ica, coop, stordalen, norgesgruppen
Topic 55	Oil price	8118	dollar, oil, barrel, brokerage, first, analyst, opec, analysts, analyst, fell, steps, securities
Topic 56	Energy	11634	energy, emissions, tons, industry, statkraft, elkem, production, aluminum, cent
Topic 57	Savings banks	5535	loss, savings, focus, kreditkassen, lost, middle, lending, positive term, bank manager
Topic 58	Expertise	3411	leader, experience, often, organization, create, people, experience, challenge, thinking
Topic 59	Offshore	8418	contract, shipyards, supply, contract, signed, offshore, building, siem, equipment, kongsberg

Topic	Label	# of articles	First words
Topic 60	Institutional investing	5241	fund, investor, investment, returns, investing, risk, managing, capital, place, private oil fund
Topic 61	Russia	6155	russia, west, russian, east, poland, moscow, president, soviet union, ukraine, authorities
Topic 62	Education	6675	school, university, professor, student, educate, research, studies, subjects, bi, institute
Topic 63	Health care	5034	hospital, physician, health, patient, human, treatment, medicine, help, expensive, develop
Topic 64	Shareholders	5456	orkla, shareholder, chairman, competition, general, bankruptcy, creditor, investor
Topic 65	Macroeconomics	8457	economy, unemployment, lower, forecast, economist, consumption, high, demand
Topic 66	Housing	11098	housing, property, real estate, apartment, square, houses, condos, land, rent, move
Topic 67	Government regulations	3360	rules, government, competition, regulations, prohibitions, competition authorities
Topic 68	Results	1545	number, growth, average, proportion, increase, decrease, compare, roughly, city
Topic 69	Publishing	4425	publishing, books, book, gyldendal, cappelen, smith, aschehoug, book club, copy, nygaard
Topic 70	Norwegian politics	11054	party, right, ap, labour, stoltenberg, political, frp, sv, election, parliamentary, politics, left
Topic 71	Norwegian counties	8035	municipality, trondheim, north, tromsø, nsb, county, local, municipal, kristiansand
Topic 72	Taxation	5171	tax, income tax, wealth tax, property, remove, lower paid, amount, system, compute
Topic 73	Quarterly results	9746	quarter, deficit, surplus, operating, tax, third one, half, group, fourth, minus, last year
Topic 74	Unknown	557	him, took, did, never, later, began, stood, gave, name, old, man, did, thought, happened
Topic 75	Entertainment	11182	film, music, record, play, artist, movie, cd, band, singing, playing, public, record, scene
Topic 76	Labor unions	8077	lo, nho, members, pay, union, strike, organization, settlement, los, union, valla, settlements
Topic 77	Fear	847	locked, fear, frame, cutting, crisis, hard, seriously, lose, dramatically, worst, consequence
Topic 78	Private banking	5975	dnb, storebrand, north, merger, bud, mutual insurance, uni, insurers, shareholders
Topic 79	Public debate	2617	political, society, debate, power, politics, politician, politicians, public, system, roll

However, it should be said labeling plays a very little role here. It is merely a convenient way of referring to the different topics.

Define T as the number of topics and let

$$P(w_i) = \sum_{j=1}^{T} P(w_i|z_i = j)P(z_i = j),$$

be the probability of word i occurring in a given document. As such, w_i is word i and z_i is a latent variable denoting which topic word i is drawn from. Different documents come with different topic distributions. D is the number of documents in the corpus and W is the number of unique words. We represent the importance of the words for different topics as the conditional probability

$$P(w_i|z = j) = \phi_w^{(j)}, \quad \text{for all } j \in [1, T] \text{ and } w_i \in \{w_1, w_2, ..., w_W\}.$$

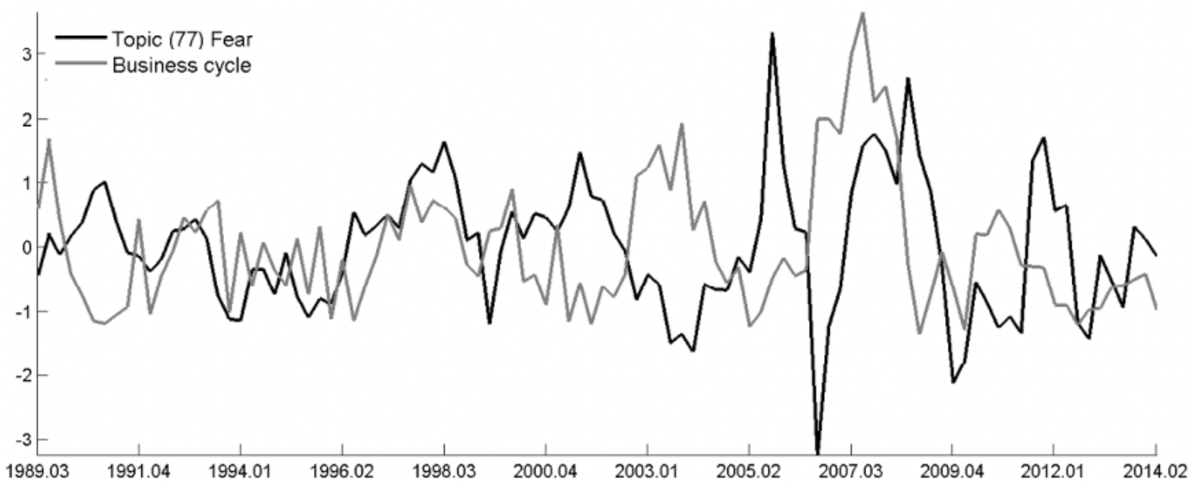
Here ϕ is a set of T multinomial distributions over the W words. The importance of a topic within a given document is defined as

$$P(z = j) = \theta_j^{(d)}, \quad \text{for all } j \in [1, T] \text{ and } d_i \in \{d_1, d_2, ..., d_D\}.$$

θ is a set of D multinomial distributions over the T topics. Given D, T, and W, the goal is to obtain estimates of ϕ and θ that maximizes the probability that a word appears in the corpus.

News index

They translate the decomposition of the DN corpus into time series. These can be used to evaluate the value of news in explaining economic fluctuations. The frequency with which each topic is represented in the newspaper is calculated every day. In the long run, each topic will have relatively equal probability of being represented in the newspaper. However, there can be substantial variation over the short run (i.e., months or quarters).



They run a battery of predictive regressions in order to investigate the predictive poewr of the news topics. The outcome variables they consider are output (Y), investment (I), consumption (C), total factor productivity (TFP), asset prices ($OSEBX$) and business confidence (BCI). For each output, two regressions are ran: one with topic included (ARX) and one without topic (AP). All models are estimated using a so-called Latent Threshold Model (LTM). The LTM can be written as

$$y_t = x'_{t-1}b_t + u_t \quad u_t \sim N(0, \sigma_u^2) b_t = \beta_t \zeta_t \quad \zeta_t = I(|\beta_t| \geq d)\beta_t = \beta_{t-1} + e_t \quad e_t \sim N(0, \Sigma_e)$$

t is the time index, x_{t-1} is a $(n \times 1)$ vector of (lagged) variables used for prediction, with b_t being a $(n \times 1)$ vector of time-varying parameters.

They focus on news predicting asset prices, i.e. $OSEBX$. They construct the aggregated news index, denoted NI_t, for each time t, based on a weighting formula:

$$NI_t = \sum_{i=1}^{T} w_i b_{i,t} n_{i,t-1},$$

where $n_{i,t-1}$ is topic i at time $t-1$. $b_{i,t}$ is the estimated parameter for topic i at time t. w_i is the weight attached to topic i in predicting y_t. The w_i are constructed using the marginal likelihoods from each $ARX(p)$:

$$w_i = \frac{p(y|M_i)}{\sum_{i=1}^{T} p(y|M_i)}.$$

Although they have weighted the topics according to how well they predict asset prices, the aggregated news index does not resemble asset prices perfectly.