

This page is the results for comparison between our semantic dependent word pair generative model (SDWP) and unigram LDA.

1. Data

Sentence is represented by semantic dependent pairs of noun and adjective in our project. A noun vocabulary and adjective vocabulary are obtained for SDWP.

For comparison, sentence based unigram LDA is employed on the noun vocabulary and non-discriminating mixture of noun and adjective by merging the noun and adjective vocabularies.

We conduct the experiments on two public datasets.

- "CitySearch" dataset contains user reviews of restaurants in US, which is crawled from newyork.citysearch.com.































- "TripAdvisor" dataset includes user reviews on hotels, which is crawled from tripadvisor.com.

Preprocessing steps include first Part-of-speech (POS) tagging with OpenNLP packages, and then dependency parsing by Malt Parser.

Then a standard stop-word list is used and dozens of corpus frequent terms is removed. After removal of nouns, neighbors and sentences with only one occurrence, a noun vocabulary of 3377 noun fragments and a dictionary adjective vocabulary of 2583 adjective neighbors are obtained from 16456 sentences for CitySearch.

The details of the two datasets are shown in Table 1.  

Table 1. Summary of the dataset used in the experiment

Item Product  	#Sentence  	#Word Pair  	  Size of noun vocabulary	  Size of neighbor (adjective) vocabulary
Restaurant  	16456  	38751  	3377  	2853  
Hotel  	16914  	44022  	4024  	3236  

2. Experimental results

The 15 clusters extracted by SDWP and unigram LDA from CitySearch and TripAdvisor dataset are shown in Table 2 and Table 3.

Table 2. The 15 clusters extracted by SDWP and unigram LDA from City Search dataset

1	Noun Adjective	<p>tables, space, spot, seating, table, dining room, bar, dining, party, cafe</p> <p>outdoor, comfortable, empty, hot, main, cozy, outside, tiny, east, private</p>
2	Noun Adjective	<p>chicken, steak, tuna, salmon, duck, oil, shrimp, bass, appetizer, gras</p> <p>grilled, delicious, fried, perfect, amazing, dry, olive, rare, tasty, special</p>
3	Noun Adjective	<p>staff, waiter, waitstaff, waitress, wait staff, waiters, server, servers, bartender, bartenders</p> <p>friendly, attentive, helpful, knowledgeable, polite, pleasant, professional, courteous, knowledgable, cute</p>
4	Noun Adjective	<p>experience, meal, night, dinner, service, lunch, visit, spot, waiter, dining experience</p> <p>overall, real, late, perfect, special, romantic, amazing, entire, quiet, pleasant</p>
5	Noun Adjective	<p>atmosphere, deco, ambience, ambiance, space, crowd, service, setting, vibe, environment</p> <p>cozy, warm, romantic, cool, beautiful, perfect, comfortable, friendly, trendy, modern</p>
6	Noun Adjective	<p>desserts, wine, chocolate, drinks, dessert, cake, hot dogs, sangria, margaritas, brulee</p> <p>delicious, hot, tasty, white, perfect, sweet, amazing, complimentary, creme, rich</p>
7	Noun Adjective	<p>music, lighting, decor, interior, walls, wood, space, dining room, band, tables</p> <p>live, white, loud, warm, soft, low, cool, dark, beautiful, bright</p>

8	Noun Adjective	<u>potatoes</u> , <u>fries</u> , <u>toast</u> , <u>cheese</u> , <u>chicken</u> , <u>spinach</u> , <u>bread</u> , <u>burgers</u> , <u>potato fries</u> , <u>appetizers</u> <u>delicious</u> , <u>french</u> , <u>sweet</u> , <u>mashed</u> , <u>fried</u> , <u>grilled</u> , cold, <u>tasty</u> , soft, blue
9	Noun Adjective	<u>bread</u> , <u>crust</u> , <u>pizza</u> , <u>cheese</u> , tea, <u>pizza</u> , <u>pasta</u> , <u>sauce</u> , <u>burger</u> , <u>sausage</u> <u>thin</u> , <u>delicious</u> , <u>hot</u> , <u>cold</u> , <u>sweet</u> , <u>iced</u> , <u>warm</u> , amazing, <u>thick</u> , <u>fantastic</u>
10	Noun Adjective	<u>dishes</u> , menu, <u>course</u> , <u>fish</u> , <u>sushi</u> , <u>courses</u> , <u>dish</u> , <u>appetizers</u> , <u>rolls</u> , <u>salads</u> main, <u>delicious</u> , <u>tasty</u> , <u>fixe</u> , <u>special</u> , <u>creative</u> , <u>unique</u> , amazing, simple, <u>raw</u>
11	Noun Adjective	<u>prices</u> , wine list, <u>drinks</u> , service, <u>menu</u> , <u>price</u> , <u>wines</u> , <u>wine</u> , <u>cocktails</u> , desserts <u>reasonable</u> , delicious, amazing, <u>extensive</u> , <u>affordable</u> , <u>worth</u> , <u>expensive</u> , <u>tasty</u> , <u>average</u> , <u>priced</u>
12	Noun Adjective	<u>ribs</u> , rice, <u>fish</u> , <u>meat</u> , <u>rib</u> , <u>beans</u> , <u>beef</u> , <u>chicken</u> , <u>pork</u> , <u>lamb</u> <u>delicious</u> , <u>fried</u> , short, tender, black, <u>flavorful</u> , <u>grilled</u> , <u>dry</u> , <u>tasty</u> , <u>sweet</u>
13	Noun Adjective	<u>service</u> , <u>ambience</u> , <u>services</u> , <u>attentive</u> , wine, <u>seating</u> , <u>attitude</u> , <u>ambiance</u> , meal, <u>staffs</u> <u>friendly</u> , <u>attentive</u> , <u>prompt</u> , <u>efficient</u> , <u>slow</u> , <u>warm</u> , professional, <u>impeccable</u> , <u>quick</u> , <u>helpful</u>
14	Noun Adjective	<u>menu</u> , <u>cuisine</u> , <u>dishes</u> , <u>restaurants</u> , <u>fare</u> , <u>flavors</u> , <u>ingredients</u> , <u>cooking</u> , wines, <u>bistro</u> <u>italian</u> , <u>french</u> , <u>japanese</u> , <u>authentic</u> , <u>traditional</u> , <u>asian</u> , <u>american</u> , <u>mexican</u> , <u>indian</u> , delicious
15	Noun Adjective	<u>service</u> , <u>waiter</u> , <u>staff</u> , <u>hostess</u> , <u>waitress</u> , <u>manager</u> , <u>waiters</u> , drinks, <u>management</u> , atmosphere <u>rude</u> , <u>bad</u> , <u>slow</u> , <u>poor</u> , <u>terrible</u> , <u>horrible</u> , <u>worst</u> , <u>awful</u> , <u>obnoxious</u> , <u>cold</u>

(a) Top 10 terms of each cluster from the 15 clusters extracted by SDWP on CitySearch dataset. Each cluster contains 10 representative nouns and the most associated 10 adjectives. Underline indicates manually evaluated semantic relevance to the cluster.

1	<u>steak</u> , <u>chicken</u> , <u>ribs</u> , <u>potatoes</u> , <u>shrimp</u> , <u>fries</u> , <u>tuna</u> , <u>spinach</u> , <u>salmon</u> , sandwich
2	<u>chocolate</u> , <u>desserts</u> , <u>cupcakes</u> , <u>gras</u> , <u>cookies</u> , <u>specials</u> , <u>cakes</u> , <u>crepes</u> , <u>pies</u> , <u>pastries</u>
3	<u>service</u> , <u>staff</u> , <u>atmosphere</u> , decor, prices, <u>ambiance</u> , <u>waitstaff</u> , menu, experience, <u>music</u>
4	<u>staff</u> , <u>atmosphere</u> , drinks, <u>crowd</u> , <u>music</u> , bar, dining room, night, spot, <u>bartenders</u>
5	staff, dishes, <u>wine list</u> , <u>menu</u> , <u>wines</u> , <u>wine</u> , owners, <u>tapas</u> , <u>price</u> , <u>bartender</u>
6	<u>tables</u> , <u>decor</u> , <u>space</u> , <u>interior</u> , <u>lighting</u> , <u>walls</u> , bar, <u>dining room</u> , <u>atmosphere</u> , wood
7	<u>pizza</u> , <u>bread</u> , <u>oil</u> , <u>crust</u> , <u>fare</u> , <u>course</u> , <u>portion</u> , <u>steaks</u> , <u>ingredients</u> , <u>fish</u>
8	<u>cheese</u> , <u>toast</u> , <u>flavor</u> , <u>burger</u> , <u>bread</u> , style, dish, <u>potato fries</u> , <u>salad</u> , <u>hot dog</u>
9	<u>chicken</u> , menu, <u>beef</u> , sauce, <u>meat</u> , rice, <u>rib</u> , <u>mussels</u> , dumplings, noodles
10	<u>service</u> , <u>waitress</u> , <u>waiter</u> , experience, <u>waiters</u> , <u>servers</u> , <u>hostess</u> , <u>server</u> , night, <u>wait staff</u>
11	<u>dishes</u> , <u>flavors</u> , <u>sushi</u> , <u>menu</u> , <u>cuisine</u> , <u>fish</u> , service, <u>rolls</u> , <u>chef</u> , fare
12	<u>coffee</u> , <u>tea</u> , meats, <u>bistro</u> , <u>ice cream</u> , <u>pie</u> , Zimbabwe, <u>baristas</u> , <u>juggle orders</u> , home, <u>cheeses</u>
13	<u>margaritas</u> , dinner, <u>brunch</u> , <u>cafes</u> , afternoon armchair, armchair, <u>cream</u> , <u>guacamole</u> , services, saturday night
14	<u>service</u> , <u>restaurants</u> , <u>prices</u> , <u>spot</u> , <u>meals</u> , party, addition, trip, <u>patio</u> , joint
15	<u>appetizers</u> , <u>courses</u> , <u>entrees</u> , chicken, <u>dessert</u> , salmon, potatoes, <u>appetizer</u> , <u>course</u> , steak

(b) Top 10 terms of each cluster from the 15 clusters extracted by sentence based unigram LDA on the noun vocabulary of CitySearch Dataset. Underline indicates the manually evaluated semantic relevance to the cluster.



1	<u>white</u> , <u>lighting</u> , <u>low</u> , <u>decor</u> , walls, <u>warm</u> , <u>dark</u> , key, <u>tables</u> , wood
2	<u>menu</u> , <u>french</u> , <u>dishes</u> , <u>italian</u> , wine list, <u>traditional</u> , <u>asian</u> , wines, <u>cuisine</u> , <u>fare</u>
3	<u>hot</u> , <u>cold</u> , <u>chocolate</u> , spot, creme, <u>tea</u> , <u>available</u> , <u>iced</u> , <u>fantastic</u> , <u>warm</u>

4	<u>service</u> , <u>rude</u> , <u>slow</u> , <u>bad</u> , <u>waiter</u> , <u>poor</u> , <u>staff</u> , <u>worst</u> , experience, <u>horrible</u>
5	real, <u>restaurants</u> , <u>italian</u> , true, <u>mexican</u> , <u>authentic</u> , <u>culinary</u> , <u>japanese</u> , <u>chef</u> , <u>cuisine</u>
6	<u>delicious</u> , <u>menu</u> , <u>special</u> , <u>drinks</u> , <u>dishes</u> , <u>tasty</u> , <u>sushi</u> , <u>amazing</u> , <u>creative</u> , <u>fish</u>
7	<u>atmosphere</u> , <u>decor</u> , <u>cozy</u> , service, <u>warm</u> , <u>ambiance</u> , <u>beautiful</u> , cool, <u>romantic</u> , <u>ambiance</u>
8	<u>outdoor</u> , <u>seating</u> , <u>tables</u> , <u>space</u> , <u>night</u> , bar, dinner, <u>table</u> , late, beautiful
9	<u>friendly</u> , <u>service</u> , <u>staff</u> , <u>attentive</u> , <u>helpful</u> , <u>knowledgeable</u> , <u>waiter</u> , <u>prompt</u> , <u>professional</u> , <u>waitstaff</u>
10	<u>music</u> , <u>live</u> , <u>loud</u> , <u>crowd</u> , cuban, <u>mixed</u> , <u>cool</u> , <u>band</u> , <u>happy</u> , <u>jazz</u>
11	<u>prices</u> , <u>service</u> , <u>reasonable</u> , <u>experience</u> , <u>worth</u> , <u>price</u> , overall, meal, <u>decent</u> , fine
12	<u>grilled</u> , <u>delicious</u> , <u>chicken</u> , <u>fried</u> , <u>ribs</u> , sweet, short, <u>potatoes</u> , black, tender
13	<u>main</u> , <u>course</u> , <u>courses</u> , <u>dishes</u> , <u>dish</u> , east, upper, dim, single, decent
14	<u>delicious</u> , <u>fried</u> , <u>thin</u> , <u>bread</u> , <u>dry</u> , cold, <u>cheese</u> , <u>french</u> , <u>tasteless</u> , perfect
15	<u>delicious</u> , <u>sweet</u> , perfect, <u>desserts</u> , <u>wine</u> , <u>tasty</u> , <u>appetizers</u> , <u>dessert</u> , amazing, <u>margaritas</u>

(c) Top 10 terms of each cluster from the 15 clusters extracted by sentence based unigram LDA on the non-discriminating mixture of nouns and adjectives of the CitySearch dataset. Underline indicates the manual evaluated semantic relevance to the cluster.

Table 3. The 15 clusters extracted by SDWP and unigram LDA from TripAdvisor dataset

❖❖1	Noun	decor, walls, furniture, floors, elevator, bathroom, furnishings, building, lift, doors
	Adjective	modern, beautiful, clean, dirty, wooden, dark, tiny, thin, white, italian
❖❖2	Noun	stay, experience, night, noise, trip, weather, visit, view, tour, shower
	Adjective	pleasant, special, enjoyable, short, real, entire, worth, warm, due, fantastic
❖❖3	Noun	suite, courtyard, view, views, terrace, lobby, breakfast room, bar, balcony, pool
	Adjective	beautiful, top, clean, private, outdoor, fantastic, quiet, junior, comfortable,

		gorgeous??
??4	Noun	access, internet, connection, internet access, mini bar, water, drinks, photos, minibar, computer
	Adjective	free, internet, available, complimentary, easy, wireless, useful, expensive, safe, stocked??
??5	Noun	street, neighborhood, city, rome, streets, road, building, sights, sites, neighbourhood
	Adjective	quiet, busy, safe, beautiful, noisy, narrow, residential, ancient, historic, clean
??6	Noun	food, restaurants, restaurant, price, prices, hotels, bar, meal, taxis, breakfast
	Adjective	reasonable, expensive, local, italian, cheap, fantastic, amazing, average, decent, real
??7	Noun	service, staff, check, food, breakfast, value, experience, touch, room service, comments
	Adjective	friendly, helpful, efficient, personal, impeccable, clean, exceptional, poor, fantastic, quick
??8	Noun	walk, steps, walking distance, location, buses, metro, rome, bus, minutes, station
	Adjective	short, easy, spanish, close, local, cheap, central, nearest, underground, convenient
??9	Noun	fruit, coffee, breakfast, juice, meats, eggs, food, pastries, water, cuts
	Adjective	fresh, cold, available, orange, hard, delicious, sweet, italian, plentiful, warm??
??10	Noun	hotel staff, reception staff, desk staff, concierge, reception, receptionist, personnel, receptionists, gentleman, waiters
	Adjective	helpful, friendly, polite, pleasant, courteous, efficient, professional, warm, NOT helpful, happy??
??11	Noun	bed, beds, towels, pillows, screen tv, tv, sheets, accommodation, channels, linen
	Adjective	comfortable, clean, flat, hard, single, huge, comfy, fresh, firm, twin??
??12	Noun	bathroom, bathrooms, size, shower, bath, bedroom, hotel room, air, standards, hotel rooms
	Adjective	clean, modern, spacious, comfortable, adequate, huge, tiny, european, decent, fine??

◆◆13	Noun	location, location of hotel, air conditioning, windows, breakfast, view, position, spot, base, hotel location
	Adjective	perfect, ideal, central, convenient, quiet, fantastic, close, superb, clean, safe
◆◆14	Noun	breakfast, standards, breakfast buffet, breakfasts, fare, buffet, buffet breakfast, selection, style, food
	Adjective	continental, european, adequate, delicious, plentiful, american, typical, fine, italian, decent◆◆
◆◆15	Noun	staff, staffs, wait staff, mario, concierge, driver, concierges, smile, desk clerks, desk people
	Adjective	helpful, friendly, pleasant, courteous, polite, professional, attentive, efficient, welcoming, willing◆◆

(a) Top 10 terms of each cluster from the 15 clusters extracted by SDWP from the TripAdvisor dataset. Each cluster contains 10 representative nouns and the most associated 10 adjectives (neighbors).

◆◆1	◆◆experience, sight, barbara, hosts, relief, services, donato, francesca, guide, grounds
◆◆2	◆◆streets, euro, tickets, ticket, transfer, bus tour, trattorias, rating, retreat, vacation
◆◆3	◆◆access, connection, internet, computer, parking, hotel lobby, safe, bonus, supermarket, wood floors
◆◆4	◆◆staff, price, hotels, taxi, city, accommodation, taxis, driver, car, night
◆◆5	◆◆lift, door, entrance, lobby, floors, building, elevator, walls, doors, hallways
◆◆6	◆◆city, rome, location, staff, photos, view, stay, walk, experience, oasis
◆◆7	◆◆courtyard, view, windows, street, noise, bed, balcony, air, terrace, suite
◆◆8	◆◆beds, bed, pillows, towels, tv, channels, sheets, screen tv, complaint, linen
◆◆9	◆◆staff, desk staff, concierge, reception staff, reception, hotel staff, lobby, service, internet access, tour,
◆◆10	◆◆staff, food, restaurant, bar, service, breakfast, restaurants, views, drinks, meal
◆◆11	◆◆location, walk, street, restaurants, neighborhood, steps, rome, walking distance, city, buses
	◆◆staff, service, breakfast, food, atmosphere, prices, price, restaurant,

12	breakfast buffet, decor
13	breakfast, coffee, fruit, meats, eggs, juice, pastries, cuts, breakfast room, croissants
14	bathroom, bed, shower, beds, size, standards, towels, bedroom, decor, bath
15	staff, breakfast, location, hotel staff, stay, standards, bathrooms, service, hotels, access

(b) Top 10 terms of each cluster from the 15 clusters extracted by sentence based unigram LDA on the noun vocabulary of TripAdvisor dataset.

1	personal, warm, late, service, night, efficient, check, due, welcome, weathe
2	tiny, windows, noise, thin, shower, noisy, walls, loud, air conditioning, minor
3	clean, towels, fresh, white, beautiful, modern, bathroom, decor, furniture, floors
4	staff, stay, rude, NOT helpful, enjoyable, experience, overall, pleasant, special, entire,
5	beautiful, view, location, top, views, perfect, fantastic, amazing, courtyard, private
6	food, reasonable, restaurant, expensive, service, local, price, cheap, restaurants, italian
7	european, standards, italian, hotels, american, typical, size, traditional, standard, average
8	quiet, location, street, busy, neighborhood, safe, city, perfect, beautiful, restaurants
9	free, internet, access, available, internet, complimentary, useful, wireless, water, connectio
10	fresh, cold, breakfast, fruit, coffee, eggs, juice, meats, orange, pastries
11	location, short, easy, walk, spanish, perfect, central, close, steps, convenient
12	bed, comfortable, beds, flat, hard, single, comfy, screen tv, twin, clean
13	breakfast, continental, delicious, adequate, plentiful, typical, fresh, breakfasts, cold, tasty,
14	clean, bathroom, comfortable, modern, spacious, bathrooms, shower, size, huge, beautiful

15	staff, helpful, friendly, pleasant, polite, courteous, hotel staff, efficient, professional, attentive
----	--------------------------------------------------------------------------------------------------------

(c) Top 10 terms of each cluster from the 15 clusters extracted by sentence based unigram LDA on the non-discriminating mixture of nouns and adjectives of TripAdvisor dataset.