Thesis Meeting 5

kense, for the thesis

Tasks

- Obtain and label sample data

- Apply distance function on sample data

- Consider automatic labeling

Labeling

- Concerns for label categories for different contexts -> sub-conversations

- New labeling:

    - **recall** - utterance that includes quoted speech or a recall to an event that occurs in the past. (i.e: "He said, "Don't you remember me?" and I was so confused.")

    - **misc** - any action or gesture that doesn't fall under utterance. (i.e: singing)

    - **give.statement** - value neutral utterance meant to give information (i.e: it's 5:00am)

    - **relax.atmosphere** - now includes exclamations (i.e: "Let's go!")

Current Labels (11)

open.question - a question that expects more than an exact singular response.

closed.question - a question that expects an exact singular response.

respond.agree - response to posed question that is in the affirmative or satisfies the question parameters.

respond.deny - response to the posed question that is in the negative, or denies answering the question.

give.statement - a value neutral utterance meant to provide information, includes commands.

give.opinion - an opinion given to the other speaker on the topic.

deflection - a response that is neither an affirmative/negative, and satisfies the parameters of the question, but the other speaker cannot confirm whether this satisfaction is valid or has occurred.

use.social.convention - speaker engages in a social convention (i.e: introduction, polite nod, greeting, gestures)

relax.atmosphere - speaker engages in a social convention with the intent to affect the mood (i.e: laughter, inside joke/statement based on rapport), *includes exclamations*.

recall - utterance that includes quoted speech or a recall to an event that occurs in the past.

misc - any action or gesture that doesn't fall under utterance.

U	L
Today is the debut of YG's newest girl group in seven years.	use.social.convention
Introducing BLACKPINK!	use.social.convention
â™ª BLACKPINKâ€¦ â™ª	use.social.convention
Please give it up for BLACKPINK!	use.social.convention
Hello, we're BLACKPINK!	use.social.convention
Our next guests are the best-charting	use.social.convention
female Korean group of all time.	use.social.convention
â€¦ wildly popular K-pop groupâ€¦	use.social.convention
K-pop sensations, BLACKPINK!	use.social.convention
â™ª Hit you with that ddu-du ddu-du du â™ª	misc
BLACKPINK has become	recall
the highest charting K-pop girl group ever.	recall
They also made history as they became	recall
the first K-pop girl group to perform at Coachella.	recall
Performing it live for the first time ever on US television,	use.social.convention
here is BLACKPINK!	use.social.convention
- BLACKPINK! - BLACKPINK!	misc
Congratulations, BLACKPINK.	use.social.convention
BLACKPINK!	misc
â™ª Hit you with that ddu-du ddu-du du â™ª	misc
- I'm so hungry. - Are you?	closed.question
#NAME?	x
There's a dessert. It's like a croissantÂ with injeolmi.	deflection
Ah!	relax.atmosphere
You guys don't like injeolmi?	closed.question
It tastes really good.	give.opinion
Says the Thai girl in the car.	deflection
Every time a new album comes, we're just excited. We're like,	give.opinion
"Ah, yay. Finally, one more," you know?	closed.question

Distance Algorithm

- Substitution cost

- InDel cost

- Clustering

```
In [19]: print(sub_cost(test1+test2))
```

{'recall, closed.question': 3.523190915023277, 'recall, give.opinion': 3.8993787254190617, 'recall, use.social.convention': 0.3392726415014651, 'recall, relax.atmosphere': 2.594831926773112, 'recall, respond.deny': -inf, 'recall, deflection': -inf, 'recall, misc': 0.24328603503162152, 'recall, give.statement': 2.3770776437505035, 'recall, respond.agree': 2.4243417895982877, 'recall, open.question': 4.158513172676578, 'closed.question, recall': 2.148795400241779, 'closed.question, give.opinion': 5.2179210354519245, 'closed.question, use.social.convention': 3.261431135587172, 'closed.question, relax.atmosphere': 6.502343644894418, 'closed.question, respond.deny': 7.836851393927039, 'closed.question, deflection': 8.5233519211102258, 'closed.question, misc': -inf, 'closed.question, give.statement': 6.133650072118235, 'closed.question, respond.agree': 7.40245668965801, 'closed.question, open.question': 7.546335239111096, 'give.opinion, recall': 3.708822391204619, 'give.opinion, closed.question': 5.604013837690056, 'give.opinion, use.social.convention': 2.5584237207162177, 'give.opinion, relax.atmosphere': 5.541474465144778, 'give.opinion, respond.deny': 6.204233306947483, 'give.opinion, deflection': -inf, 'give.opinion, misc': -2.9038850999994414, 'give.opinion, give.statement': 5.030832390271206, 'give.opinion, respond.agree': 5.812511939273692, 'give.opinion, open.question': 6.411611583814668, 'use.social.convention, recall': -0.07576485777737878, 'use.social.convention, closed.question': 4.061132485101341, 'use.social.convention, give.opinion': 3.9573415546317845, 'use.social.convention, relax.atmosphere': 3.142045347792994, 'use.social.convention, respond.deny': -inf, 'use.social.convention, deflection': -inf, 'use.social.convention, misc': -0.36654425409407687, 'use.social.convention, give.statement': 3.114673947748971, 'use.social.convention, respond.agree': 4.140834337417307, 'use.social.convention, open.question': 3.126758998686796, 'relax.atmosphere, recall': 1.767668523635328, 'relax.atmosphere, closed.question': 5.5140729374765245, 'relax.atmosphere, give.opinion': 5.3327470170322195, 'relax.atmosphere, use.social.convention': 4.784088943721418, 'relax.atmosphere, respond.deny': 3.256415709097181, 'relax.atmosphere, deflection': -inf, 'relax.atmosphere, misc': -inf, 'relax.atmosphere, give.statement': 5.3735083650053479, 'relax.atmosphere, respond.agree': 5.816784356462005, 'relax.atmosphere, open.question': 6.429846654041029, 'respond.deny, recall': -inf, 'respond.deny, closed.question': 7.421813894648195, 'respond.deny, give.opinion': 3.9818408856110348, 'respond.deny, use.social.convention': 1.6122911359078806, 'respond.deny, relax.atmosphere': 3.256415709097181, 'respond.deny, deflection': -inf, 'respond.deny, misc': -inf, 'respond.deny, give.statement': 7.199869210415775, 'respond.deny, respond.agree': 2.9846422360851745, 'respond.deny, open.question': -inf, 'deflection, recall': 6.771118428466863, 'deflection, closed.question': -inf, 'deflection, give.opinion': 7.405307006960459, 'deflection, use.social.convention': -inf, 'deflection, relax.atmosphere': 7.527878737001556, 'deflection, respond.deny': -inf, 'deflection, misc': -inf, 'deflection, give.statement': -inf, 'deflection, respond.agree': -inf, 'deflection, open.question': -inf, 'misc, recall': -inf, 'misc, closed.question': -inf, 'misc, give.opinion': 0.34404241344414394, 'misc, use.social.convention': 0.8350896070755736, 'misc, relax.atmosphere': -2.044347775792139, 'misc, respond.deny': -inf, 'misc, deflection': -inf, 'misc, give.statement': -0.981312658720872, 'misc, respond.agree': -inf, 'misc, open.question': -inf, 'give.statement, recall': 3.4992247530362937, 'give.statement, closed.question': 7.201354931256626, 'give.statement, give.opinion': 4.79878596271644, 'give.statement, use.social.convention': 3.9305908833099994, 'give.statement, relax.atmosphere': 5.452285478437597, 'give.statement, respond.deny': 6.217571212149235, 'give.statement, deflection': 7.131482235435525, 'give.statement, misc': -0.981312658720872, 'give.statement, respond.agree': 4.881249536469684, 'give.statement, open.question': 3.3744870703100656, 'respond.agree, recall': 3.0808575513444776, 'respond.agree, closed.question': 6.438590475711724, 'respond.agree, give.opinion': 5.14167160339213, 'respond.agree, use.social.convention': 2.3547741937229314, 'respond.agree, relax.atmosphere': 5.558986598994359, 'respond.agree, respond.deny': 5.569604736806331, 'respond.agree, deflection': 7.671142763268393, 'respond.agree, misc': -inf, 'respond.agree, give.statement': 6.267655943531465, 'respond.agree, open.question': 7.694126081269233, 'open.question, recall': 0.9361207513401298, 'open.question, closed.question': 6.747247933037093, 'open.question, give.opinion': 5.652103420997374, 'open.question, use.social.convention': 2.96629432649355, 'open.question, relax.atmosphere': 5.195381400404006, 'open.question, respond.deny': -inf, 'open.question, deflection': 10.480602273650922, 'open.question, misc': -inf, 'open.question, give.statement': 5.826999275007572, 'open.question, respond.agree': 7.766882423704547}

```python
print(sub_cost(test1+test3))
```

{'recall, closed.question': 3.8035271075966586, 'recall, give.opinion': 5.179544295097426, 'recall, use.social.convention': 1.7085675441562718, 'recall, relax.atmosphere': 3.2031720469263423, 'recall, respond.deny': -inf, 'recall, deflection': -inf, 'recall, x': -inf, 'recall, mis': -inf, 'recall, misc': 4.48767661125147B, 'recall, give.statement': 3.4712020622339392, 'recall, respond.agree': 3.941737049015713, 'recall, open.question': 4.7420051322944765, 'closed.question, recall': 2.3661217952893603, 'closed.question, give.opinion': 4.7206766001728715, 'closed.question, use.social.convention': 4.7278556870839, 'closed.question, relax.atmosphere': 9.065832869596106, 'closed.question, respond.deny': 10.111076517203145, 'closed.question, deflection': 11.959073423758095, 'closed.question, x': 6.4215580854394, 'closed.question, mis': -inf, 'closed.question, misc': -inf, 'closed.question, give.statement': 8.948405768388406, 'closed.question, respond.agree': 9.617278986850287, 'closed.question, open.question': 9.37967351064219, 'give.opinion, recall': 5.062998678437681, 'give.opinion, closed.question': 5.992393249106466, 'give.opinion, use.social.convention': 3.327124032845372, 'give.opinion, relax.atmosphere': 5.2739526827714185, 'give.opinion, respond.deny': 5.787899453161257, 'give.opinion, deflection': -inf, 'give.opinion, x': 5.703243079556373, 'give.opinion, mis': -inf, 'give.opinion, misc': 4.187248988379885, 'give.opinion, give.statement': 5.282505409883566, 'give.opinion, respond.agree': 4.879541426357802, 'give.opinion, open.question': 6.251667399201726, 'use.social.convention, recall': 2.941228300946547, 'use.social.convention, closed.question': 5.070743400606908, 'use.social.convention, give.opinion': 3.923643564988928, 'use.social.convention, relax.atmosphere': 4.5952795877644, 'use.social.convention, respond.deny': -inf, 'use.social.convention, deflection': -inf, 'use.social.convention, x': 3.120147644531586, 'use.social.convention, mis': -inf, 'use.social.convention, misc': 3.4586916344927343, 'use.social.convention, give.statement': 6.173875395761965, 'use.social.convention, respond.agree': 4.860189062166285, 'use.social.convention, open.question': 4.11815579685789, 'relax.atmosphere, recall': 2.006135199866789, 'relax.atmosphere, closed.question': 7.824540242916378, 'relax.atmosphere, give.opinion': 5.470989529830972, 'relax.atmosphere, use.social.convention': 6.913684616088066, 'relax.atmosphere, respond.deny': -inf, 'relax.atmosphere, deflection': -inf, 'relax.atmosphere, x': -inf, 'relax.atmosphere, mis': 10.744035164246624, 'relax.atmosphere, misc': -0.12066687350841193, 'relax.atmosphere, give.statement': 7.547391904705782, 'relax.atmosphere, respond.agree': 8.081070151524195, 'relax.atmosphere, open.question': 7.593093266171884, 'respond.deny, recall': 5.5424497832850825, 'respond.deny, closed.question': 7.789148422315782, 'respond.deny, give.opinion': 5.372861953882413, 'respond.deny, use.social.convention': -inf, 'respond.deny, relax.atmosphere': 6.6565723229962B4, 'respond.deny, deflection': -inf, 'respond.deny, x': -inf, 'respond.deny, mis': -inf, 'respond.deny, give.statement': 9.761778393236712, 'respond.deny, respond.agree': 6.664323957901043, 'respond.deny, open.question': -inf, 'deflection, recall': -inf, 'deflection, closed.question': 10.696039017924301, 'deflection, give.opinion': 3.7878994531612573, 'deflection, use.social.convention': -inf, 'deflection, relax.atmosphere': 10.563462918604802, 'deflection, respond.deny': -inf, 'deflection, x': -inf, 'deflection, mis': -inf, 'deflection, misc': -inf, 'deflection, give.statement': -inf, 'deflection, respond.agree': -inf, 'deflection, open.question': -inf, 'x, recall': 1.531003256633974, 'x, closed.question': -inf, 'x, give.opinion': 6.821757550815763, 'x, use.social.convention': -inf, 'x, relax.atmosphere': 6.5520163919536945, 'x, respond.deny': -inf, 'x, deflection': -inf, 'x, mis': -inf, 'x, misc': -inf, 'x, give.statement': -inf, 'x, respond.agree': -inf, 'x, open.question': -inf, 'mis, recall': -inf, 'mis, closed.question': -inf, 'mis, give.opinion': -inf, 'mis, use.social.convention': -inf, 'mis, relax.atmosphere': -inf, 'mis, respond.deny': -inf, 'mis, deflection': -inf, 'mis, x': -inf, 'mis, misc': -inf, 'mis, give.statement': -inf, 'mis, respond.agree': -inf, 'mis, open.question': 13.165031759320899, 'misc, recall': 3.541314574853551, 'misc, closed.question': 0.011909225811087035, 'misc, give.opinion': 3.7942042174430965, 'misc, use.social.convention': 4.6906383625424315, 'misc, relax.atmosphere': -0.12066687350841193, 'misc, respond.deny': -inf, 'misc, deflection': -inf, 'misc, x': 3.761753187876857, 'misc, mis': -inf, 'misc, give.statement': 4.569501697453172, 'misc, respond.agree': -inf, 'misc, open.question': -inf, 'give.statement, recall': 4.329473952622279, 'give.statement, closed.question': 9.495420393781481, 'give.statement, give.opinion': 4.685919025065516, 'give.statement, use.social.convention': 6.30427203301764, 'give.statement, relax.atmosphere': 7.853534836954657, 'give.statement, respond.deny': 8.868693597153223, 'give.statement, deflection': 9.231263676537932, 'give.statement, x': 3.3718202433318742, 'give.statement, mis': -inf, 'give.statement, misc': 3.0914544006485283, 'give.statement, respond.agree': 8.104482130600976, 'give.statement, open.question': 6.158324290464466, 'respond.agree, recall': 4.754366913113992, 'respond.agree, closed.question': 9.045140526486716, 'respond.agree, give.opinion': 5.3852644153343086, 'respond.agree, use.social.convention': 2.567407312938439, 'respond.agree, relax.atmosphere': 7.6001801049507245, 'respond.agree, respond.deny': 9.2492864586222, 'respond.agree, deflection': -inf, 'respond.agree, x': -inf, 'respond.agree, mis': -inf, 'respond.agree, misc': -inf, 'respond.agree, give.statement': 7.889469239630125, 'respond.agree, open.question': 9.722835425455255, 'open.question, recall': 3.1570426315733204, 'open.question, closed.question': 8.282062714015769, 'open.question, give.opinion': 5.769671477431758, 'open.question, use.social.convention': 4.11815579685789, 'open.question, relax.atmosphere': 7.233550879484744, 'open.question, respond.deny': -inf, 'open.question, deflection': 10.662531418791716, 'open.question, x': 6.388050486306813, 'open.question, mis': -inf, 'open.question, misc': 3.8852922222870196, 'open.question, give.statement': 8.311276213303065, 'open.question, respond.agree': 9.693291149841238}