Automatic Extraction of Semantic Patterns in Dialogs using Convex Polytopic Model

Jingyan Zhou[1], *Xiaoying Zhang*[1], *Xiaohan Feng*[1], *King Keung Wu*[2], *Helen Meng*[1]

[1]Dept. of Systems Engineering Engineering Management, Chinese University of Hong Kong
[2] SpeechX Limited

{jyzhou, zhangxy, xhfeng, hmmeng}@se.cuhk.edu.hk, kkwu@speechx.cn

2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP) | 978-1-7281-6994-1/20/$31.00 ©2021 IEEE | DOI: 10.1109/ISCSLP49672.2021.9362051

Abstract

Natural Language Understanding (NLU) in task-oriented dialog systems usually requires annotated data for training the understanding module. Annotation of large data sets is a costly process. This paper proposes an unsupervised framework based on Convex Polytopic Model (CPM), which automatically extracts *semantic patterns* from a raw dialog corpus using a geometric approach to assist in generating the semantic frames. We discover that the semantic patterns extracted are easily interpretable and have a strong correlation with the intent and slots of the semantic frames and may potentially serve as the basic units for NLU. This is an initial investigation of the properties of CPM to explore its semantic interpretability. Experiments are based on the ATIS (Air Travel Information System) corpora and show that CPM can generate semantic frames with minimal supervision.

Index Terms: Semantic Pattern, Convex Polytopic Model, Natural Language Understanding, Task-oriented Dialog System

1. Introduction

Task-oriented dialog systems aim to assist users in accomplishing a specific task, e.g. restaurant reservations, weather querying, flight reservations, etc.. For a dialog system serving a specific application domain, there is an underlying knowledge representation of the domain. This representation often consists of the user's possible intents and related attributes (often referred as slots with values) [1]. For example, Figure 1 shows a semantic frame of a given utterance in the ATIS (Air Travel Information Systems) domain.

The problem of Natural Language Understanding involves transforming the user's natural language input (e.g. spoken utterance or typed sentence) into the semantic frame. Conventional approaches include grammar-based parsing [2, 3], machine learning techniques such as Conditional Random Fields [4] and neural approaches [5, 6, 7], which consider intent identification as a multi-class classification problem, and slot-filling as a sequence classification problem. Grammar-based parsing calls for the design of grammar rules, and machine learning techniques call for sizeable annotated data. Both grammar design and data annotation are laborious processes that are based upon an underlying knowledge representation. Therefore, it will be desirable if the knowledge representation can be obtained somewhat automatically and efficiently.

This paper presents a novel, data-driven framework based on the **Convex Polytopic Model (CPM)** [8] to extract the key *semantic patterns* from the raw conversational data. CPM is an unsupervised geometric algorithm for automatically extracting the key concepts from texts, e.g. discovering topics from a document corpus where the documents are assumed to consist of mixtures of topics [8]. Therefore, after projecting all utterances as points onto an affine subspace via Principal Component Analysis (PCA) and enclosing all the points by a compact convex polytope, we can then interpret the vertices (i.e. extreme points) of the polytope as the representation of key semantic patterns.

We show that such a geometric representation is interpretable and indicates intent and slot information, and hence can potentially derive the basic units to support NLU. The basic units can guide the human to identify key semantic features in the corpus that relate to the knowledge representation (or ontology) of the application domain. The units can also from the basis for deriving more sophisticated semantic structures, such as defining grammar rules for parsing.

The contribution of our work is three-fold: 1) we propose an unsupervised method via CPM to automatically extract key semantic features with high interpretability from raw conversation data; 2) we investigate the two geometric properties of the output of CPM and extract semantic patterns which indicate intent and slot information; 3) we demonstrate the possibility of taking the extracted patterns as basic units for building a higher-level semantic structure for NLU in task-oriented dialog systems.

This paper is organized as follows: Section 2 introduces the detailed formulation of CPM. Section 3 presents the experiments on the ATIS corpus and verifies the properties of CPM. Section 4 concludes the paper and discusses potential future work.

What flights are there **on Sunday** **from Seattle to Chicago**

Indicate the INTENT	DATE	From ORIGIN to DESTINATION

INTENT:	*Flights*
SLOTS:	ORIGIN = *Seattle*
	DETINATION = *Chicago*
	DATE = *Sunday*

Figure 1: *An utterance from ATIS dataset querying flight information and its corresponding semantic frame.*

2. Convex Polytopic Model

This section describes the formulation of the CPM and discusses the semantic properties of its geometric characteristics on raw dialog data.

2.1. Model Description

The CPM consists of two steps. First, all utterances in the corpus are embedded in a low-dimensional affine subspace using PCA. Second, a compact convex polytope is generated to enclose all the embedded utterances points. We follow the steps of implementing CPM in [8].

Step 1: Embed utterances into a low-dimensional affine subspace using PCA

Define M as the vocabulary size of the corpus, and N as the number of utterances. The corpus is first transformed into a sum-normalized term-document matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$, where the utterance is encoded as a sum-normalized word count vector of dimension M.

After applying PCA to $\mathbf{X}$, we obtain the basis $\mathbf{U} \in \mathbb{R}^{M \times R}$. The utterances are projected onto the R-D affine subspace spanned by the basis $\mathbf{U}$. As the principal components represent the direction along which the data points have the largest variance, they can capture the semantic features which can optimally distinguish among the points.

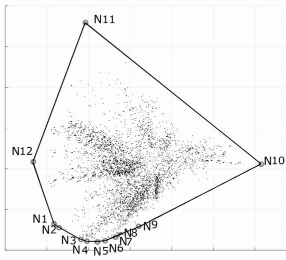
Step 2: Generate a compact convex polytope to enclose the embedded utterances

Note that we sum-normalize the utterance representation in $\mathbf{X}$. Hence the shorter utterance has fewer non-zero values, and thus the non-zero values tend to be larger due to the sum-to-one property and sparsity. Geometrically, the utterance points lie on a sum-to-one hyperplane, and those with sparser representation tend to be the extreme points (i.e. lie further away from the centre of the non-negative region of the hyperplane).

As PCA preserves the distinctive features of the utterance points, the extreme points remain extreme after PCA projection. We observe that utterances consist of combinations of semantic patterns, and those with fewer semantic patterns tend to be shorter. Hence, the utterances with a single semantic pattern generally form the extreme points (i.e. vertices), though the converse may not be true. Therefore, we use the vertices of the convex polytope with minimum volume to enclose all the utterance points and obtain the short utterances with key semantic patterns.

Note that the minimum volume convex polytope always exists for a finite point set and is unique, where the vertices are supported by some of the utterance points. This kind of polytope is referred as the **normal (NO)-type** polytope [8] and can be computed by the Quickhull algorithm [9]. The points that coincide with the vertices correspond to the shorter utterances that contain fewer semantic patterns than other utterances in the corpus.

However, our goal is to discover the semantic patterns exhibited by the utterances instead of finding the instances of the semantic patterns. Moreover, not all semantic patterns appear individually in the corpus, and hence we seek to find a better compact polytope that can disentangle and identify the distinctive, key semantic patterns. Furthermore, note that finding the NO-type convex polytope does not offer control in the number of semantic patterns. In order to allow pre-specification of the number of semantic patterns, we adopt the *simplex* method and specify the number of patterns to derive a convex polytope whose vertices can be regarded as the semantic patterns abstracted from the utterances that contain them. To make the simplex compact, we impose the constraint of minimizing the volume when generating the simplex that encloses all the points using Minimum Volume Simplex Analysis [10, 11]. Such generated simplex is referred as **Minimum Volume Simplex (MVS)** [8]. Note that vertices of the MVS-type polytope are no longer in the point set (i.e. the set of points representing the utterances). Instead, each utterance point can be uniquely represented by a convex combination of the vertices which allows better disentangling capability.



N1	from Pittsburgh to Baltimore
N2	from Philadelphia to Toronto
N3	flight from Denver to Philadelphia
N4	flight from Denver to Boston
N5	i would like to fly to Baltimore
N6	i want to find a flight from Boston to Atlanta
N7	i need a flight from Boston to Denver
N8	i would like a flight from Boston to Atlanta
N9	i need a flight to Atlanta this afternoon
N10	what is ewr
N11	show me the flights
N12	flights to Baltimore

Figure 2: *NO-type polytope and related utterances. Vertices are labelled as N1-N12. The scattered dots denote projected utterance points and the circled points are the vertices of the polytope.*

2.2. Model Properties

Based on the model description above, the CPM is expected to carry the following properties relating to the dialog corpus:

First, the utterances in the dataset are projected to an affine space and enclosed by a polytope. The utterances with a mixture of multiple semantic patterns tend to be in the interior of the polytope.

Second, the vertices (extreme points) of the polytope can capture distinct semantic patterns. Two geometric properties relate to interpretation of these patterns:

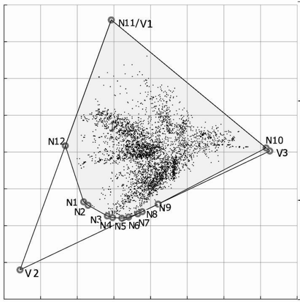
1. The **coordinates** of the vertices in the original term-document coordinate system where each dimension corresponds to a term. Hence, for each vertex, the top-k terms with k highest weights (i.e. coordinate values) are the most frequently observed terms in the corresponding pattern and hence can be used to interpret that pattern.

2. The **Euclidean distances** between the vertices and other data points. The k-nearest neighbouring data points of each vertex are the utterances that exhibit the corresponding pattern. Therefore, the semantic pattern of the vertex can be interpreted by observing the common patterns among these *k-nearest utterances*.

Third, polytopes with a higher dimensionality can preserve more diversified distinct semantic patterns. The vertices containing mixed patterns in a lower dimensional space may be disentangled into different vertices in a higher dimensional space.

3. Experiments

In this section, we verify the properties of the CPM for automatically extracting semantic patterns from a raw dialog corpus. We implemented the CPM on the same dataset as [12]: 4,978 utterances from the training sets of the ATIS-2 and ATIS-3 corpora of the air travel domain [13]. After embedding the utterances of the corpus into the low-dimensional affine subspace via PCA, we applied the Quickhull algorithm [9] to generate the NO-type convex polytope and the algorithm proposed in [10, 11] for generating the MVS-type convex polytope. The CPM experiments were implemented in Matlab.



V1	**show me** the **flights**
	show me all twa **flights** please
	show me the **flights** leaving Baltimore
	***flights**, **me**, **show**, the, all*
V2	**from** Pittsburgh **to** Baltimore
	from Philadelphia **to** Toronto
	from Denver **to** Baltimore
	***to**, **from**, i, flight, a*
V3	**what is** ewr
	what is mco
	what is dl
	***what**, **is**, in, the, ground*

Figure 3: *MVS-type polytope, related nearest 3 utterances and top 5 words (marked in italics). The region of the NO-type polytope is filled with light grey. MVS-type vertices are labelled as V1-V3. The scattered dots denote projected utterance points and the circled points are the vertices of the polytope.*

3.1. Experiment 1: Two-dimensional Analyses

3.1.1. NO-type Convex Polytope

The corpus is first transformed into a sum-normalized term-utterance matrix. Then, we projected all the utterances onto a Two-dimensional(2-D) affine subspace via PCA, and generated the NO-type polytope as shown in Figure 2, where 12 vertices are generated and all of them are within the utterance point set. The corresponding utterances of the vertices N1-N12 are shown alongside the plot.

The utterances at the vertices of the polytope are all short and contain fewer semantic patterns. The complex utterances that are combinations of multiple patterns are mapped to the interior of the polytope. For example, the utterance: "*Show me the earliest flight* on August second *from Boston to Denver* that serves a meal.", which contains multiple patterns and the patterns in italics can be traced to Vertex N11 and N1/N2 respectively, is near the centre of mass of the projected data.

We also note that there is a cluster of vertices: N1-N9 in Figure 2 as the points are close to each other (considering their Euclidean distance) and the corresponding utterances of N1-N9 also show a clear pattern: "from CITY to CITY". We use placeholders in the semantic patterns to represent the tokens of the same semantic type. Here we use "CITY" for city names, "AIRLINE" for airline names and "CODE" for abbreviations of airports and airlines in the dataset. Besides, N10, N11 and N12 show distinct semantic patterns. We expect the vertices to abstract diversified patterns from the corpus, which should be distinct from each other. Hence, MVS-type polytope is investigated in the next section for this purpose.

3.1.2. MVS-type Convex Polytope

To extract distinct semantic patterns, we generated the MVS-type polytope in 2-D subspace, and the result is shown in Figure 3, with three vertices denoted as V1, V2 and V3 respectively. As V2 and V3 are not in the utterance set, we infer the semantic pattern of each vertex by its top-k words and k-nearest utterances, which was discussed earlier in Section 2.2. Therefore, we also list the 3 nearest utterances along with the top-5 words of each vertex in Figure 3. The nearest utterances of each vertex are all short texts that have relatively simple semantic structure, and the top words are words that frequently observed in the related semantic pattern. The three vertices can be related to clear semantic patterns (marked as bold in the table shown in Figure 3). The pattern of V1 is the same as N11: "show me the/all flights" as they are located at the same position. V3 also represents "what is CODE" as it is closed to N10. If we extend the line that connects the vertices N8 and N9, as well as the line that connects N11 and N12, and then examine the intersection of the two extended lines, we observe V2, which abstracts "from CITY to CITY" from the repeating patterns N1-N9 and N12.

To study the properties of CPM further, we also have the following observations regarding the relationships among the top words of each vertex. As the PCA embedding preserves the essential features in the corpus, the corresponding top words of the three vertices frequently appear in the dataset. Besides, some of the top words in the corresponding vertex usually co-exist in the pattern (e.g. $show$, me, and $flights$, usually appear together and form the pattern $show\ me\ the/all\ flights$). Also, some of them have similar contexts (i.e. the neighbouring words), which indicates that they have similar meanings and can be categorized to the same semantic type (e.g. all and the in V1; $Philadelphia$ and $Denver$ in V2). These observations are consistent with our discussion about top words in Section 2.2.

The result above shows that the vertices of MVS-type polytope can capture distinctive semantic patterns, the process of which are interpretable through analyzing the nearest utterances and top words. However, the actual possible patterns in the corpus should be far more than three. Hence, we continue our work with MVS-type CPM on higher dimensions.

3.2. Experiment 2: Higher-Dimensional Analyses

To track the change of the patterns captured by the MVS-type CPM with the increasing number of vertices, we conduct the experiments from 2-D to 50-D (i.e. MVS-type polytope with 3 to 51 vertices). We observe that the vertices with mixed patterns in the lower-dimension polytope are split into different vertices with the dimension increased, and the polytopes with a higher dimension preserve more distinct semantic patterns. Some of them can be derived from splitting vertices in lower dimensions, while the others are newly extracted. Here, we take the 9-D (10-vertex) MVS-type convex polytope as an illustrative example. The vertices, their corresponding top-5 words and the inferred patterns are shown in Table 1.

Table 1: *Top-5 words and extracted semantic patterns of the 10 vertices generated by 9-D MVS-type convex polytope*

V	Top-5 words	Semantic patterns
V1	**me,show**,all,Baltimore,does	show me
V2	**to,from**,Boston,i,Baltimore	from CITY to CITY
V3	**what,is**,does,**fare**,mean	what is, CODE, fare code, restriction code
V4	**the,flight**,Boston,in,Atlanta	the flight
V5	**San,Francisco**,from,to,on	San Francisco
V6	Boston,**and,between**, flights, fly	between CITY and CITY
V7	**ground,transportation,in**, Dallas,Baltimore	ground transportation in CITY
V8	**Denver,Pittsburgh**,flight,to, ground	Denver, Pittsburgh
V9	**flights**,on,**from**,are,**list**	flights from CITY, list all
V10	on, **a,flight**,**i,like**	a flight, i would like/need

Comparing these results with the 2-D (3-vertex) MVS-type results, the polytope with higher dimensions preserves the same semantic patterns of V1-V3. Also, it captures new patterns, e.g. V7: "ground transportation in CITY" and V4: "the flight". Some words in the same semantic category are extracted as a distinct pattern. For example, $San\ Francisco$ at V5, and $Denver$, $Pittsburgh$ at V8 are all city names. These patterns are extracted due to many utterances containing phrases such as $CITY\ to$, $from\ CITY$, $at\ CITY$, etc. in the dataset. There are also vertices consisting of more than one patterns. For example, V9 extracts $flights\ from\ CITY$ and $list\ all$; V3 captures not only $what\ is\ CODE$ as in 2-D MVS-type convex polytope, but also $what\ is\ fare\ code$ and $what\ is\ restriction$; and V10 contains $a\ flight\ on\ AIRLINE$ and $I\ would\ like/need$. To better illustrate vertices with mixed patterns, we present the different patterns captured by V3 in Table 2 . The utterances containing these patterns and their distances from V3 are also listed (note that there are repeating utterances in the corpus).

Table 2: *Three patterns captured by V3 of 9-D MVS-type polytope. The pattern "what is" is not listed here as all of the utterances contain this pattern.*

Patterns in V3	Utterances near V3	Distance
CODE	what is ewr	0.2329
	what is mco	0.2358
restriction	what is restriction ap 57	0.3077
	what is restriction ap 80	0.3077
fare code	what is fare code h	0.3475
	what is fare code h	0.3475

As Table 2 shows, the utterances containing these three patterns are all close to V3 as they all consist of "what is ..." and one other pattern. However, with the dimension of the polytope increased to 33-D (34 vertices), all these mixed patterns are split into different vertices. As shown in Table 3, the mixed patterns in V3 of the 9-D polytope are split into V3, V27, and V30 in the new polytope separately. Also, patterns in V9 of 9-D are split to V9 and V19, while V10 is split into V10 and V29. We also listed some newly extracted patterns in Table 3, e.g. patterns indicating $Time$ information at V26: "on Wednesday/Sunday/Thursday/..." and V22: "in the morning/afternoon/evening/...".

The patterns in Table 3 contain adequate intent and slot information. Based on these patterns, we can start drafting the frames, e.g. for frame with intent *Flight* (V25), the associate slot set contains *Origin* and *Destination* (V2, V9), *Time* (V22), *Date* (V26), etc.. After labelling, these patterns can also be used for semantic frame parsing. For example, the *Intent*, *Origin*, *Destination*, and *Time* information can be extracted from the utterance in Figure 1 as it contains three patterns in Table 3: V25 "what flights", V2 "from CITY to CITY", and V26 "on Tuesday".

4. Conclusions

In this paper, we present an unsupervised data-driven framework based on CPM to facilitate the process of generating semantic frames by extracting key semantic features that contain intent and slot information in the conversational system. Our experiments also reveal that the geometric properties, including coordinates, distances and dimensions of the framework, provide high interpretability to the extracted semantic patterns.

Table 3: *Examples of extracted semantic patterns in 33-D (34 vertices) MVS-type convex polytope*

Vertex	extracted semantic patterns
V1	show me
V2	from CITY to CITY
V3	what is, CODE
V9	flights from CITY
V10	i would like/need
V19	list all
V22	in the morning/afternoon/evening/...
V25	what flights
V26	on Wednesday/Sunday/Thursday/...
V27	what is, restriction code
V29	a flight
V30	what is, fare code
...	...

This work is an initial step in investigating the Convex Polytopic Model for unsupervised semantic frame generation in conversational systems. As we have identified these patterns in the utterances, semantic frames and domain ontology can be constructed. This can contribute towards the language understanding tasks in the dialog system. We also demonstrate possible usage of applying extracted patterns to design and fill the semantic frame for NLU in dialog system. Future work will focus on automating the derivation of semantic patterns based on the geometric properties, evaluating and refining the extracted patterns, and their extension in semantic parsing to support NLU.

5. Acknowledgements

This work is partially supported by the General Research Fund from the Research Grants Council of Hong Kong SAR Government (Project No. 14245316).

6. References

[1] D. Jurafsky, *Speech & language processing*. Pearson Education India, 2000.

[2] D. Jurafsky, C. Wooters, G. Tajchman, J. Segal, A. Stolcke, E. Fosler, and N. Morgan, "The berkeley restaurant project," in *Third International Conference on Spoken Language Processing*, 1994.

[3] E. C. Kaiser, M. Johnston, and P. A. Heeman, "Profer: Predictive, robust finite-state parsing for spoken language," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)*, vol. 2. IEEE, 1999, pp. 629–632.

[4] C. Raymond and G. Riccardi, "Generative and discriminative algorithms for spoken language understanding," in *Eighth Annual Conference of the International Speech Communication Association*, 2007.

[5] J. Gao, M. Galley, and L. Li, "Neural approaches to conversational ai," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 1371–1374.

[6] C.-W. Goo, G. Gao, Y.-K. Hsu, C.-L. Huo, T.-C. Chen, K.-W. Hsu, and Y.-N. Chen, "Slot-gated modeling for joint slot filling and intent prediction," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, 2018, pp. 753–757.

[7] R. Gangadharaiah and B. Narayanaswamy, "Joint multiple intent detection and slot labeling for goal-oriented dialog," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language*

Technologies, Volume 1 (Long and Short Papers), 2019, pp. 564–569.

[8] K. K. Wu, H. Meng, and Y. Yam, "Topic discovery via convex polytopic model: A case study with small corpora," in *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 2018, pp. 000 367–000 372.

[9] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Transactions on Mathematical Software (TOMS)*, vol. 22, no. 4, pp. 469–483, 1996.

[10] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, vol. 3. IEEE, 2008, pp. III–250.

[11] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 5067–5082, 2015.

[12] D. Hakkani-Tür, G. Tür, A. Celikyilmaz, Y.-N. Chen, J. Gao, L. Deng, and Y.-Y. Wang, "Multi-domain joint semantic frame parsing using bi-directional rnn-lstm." in *Interspeech*, 2016, pp. 715–719.

[13] P. Price, "Evaluation of spoken language systems: The atis domain," in *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27, 1990*, 1990.