

ICDE2015勉強会

HaTen2: Billion-scale Tensor Decompositions

**I. Jeon (KAIST), E. E. Papalexakis (CMU), U Kang (KAIST), C.
Faloutsos (CMU)**

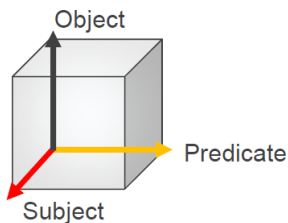
担当：丸橋（富士通研）

テンソル分解とは

図は論文またはポスターより抜粋

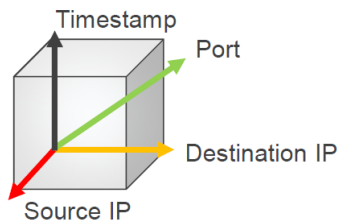
テンソルとして表現
できるデータは多い

3D Tensor



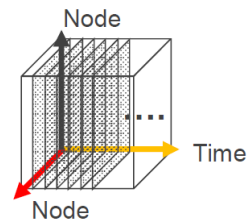
Knowledge base

4D Tensor



Network traffic logs

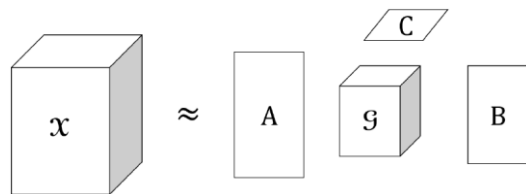
3D Tensor



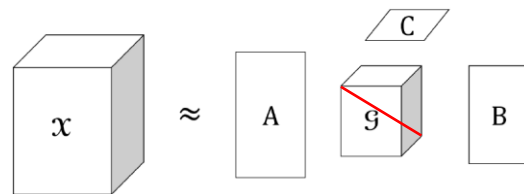
Graph streams

テンソル分解：
テンソルを行列と
コアテンソルの積
に近似

Tucker分解

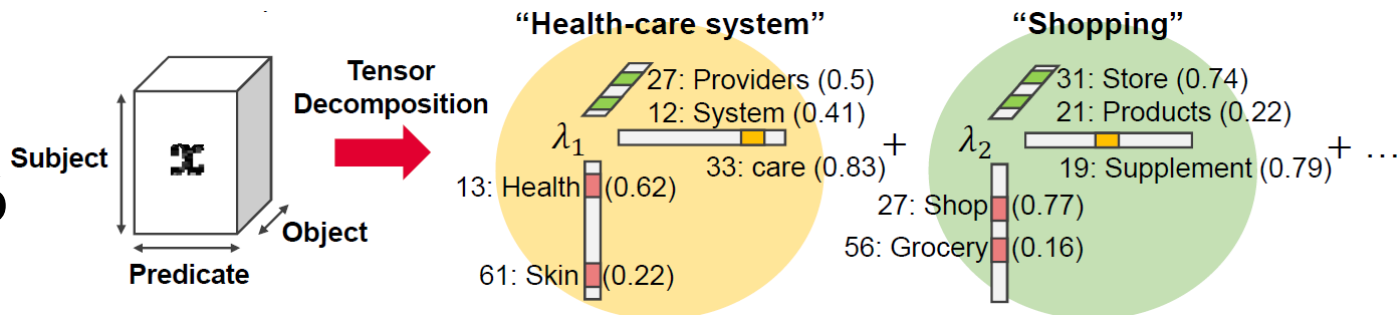


PARAFAC分解



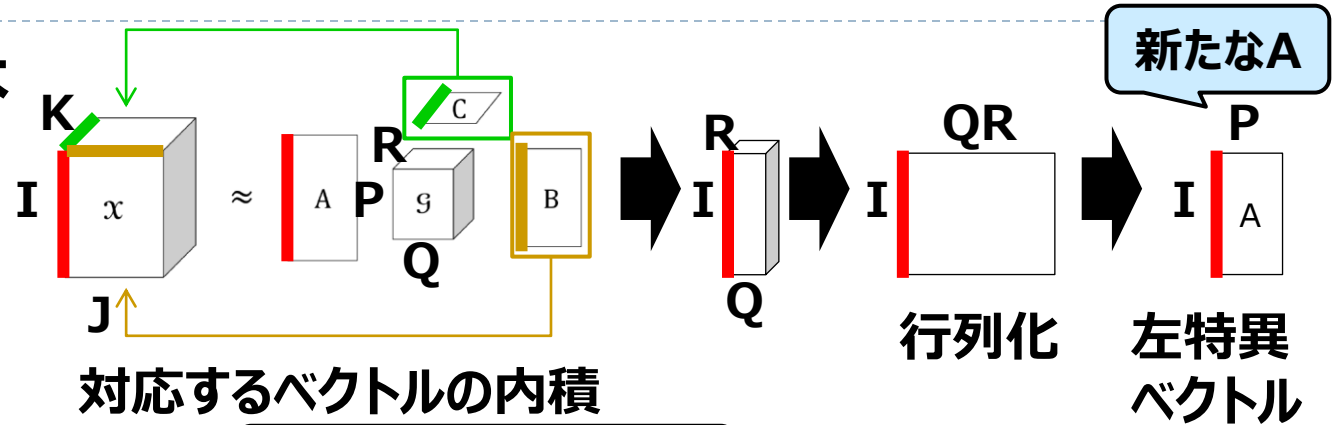
(コアテンソルが正方対角)

何が嬉しい？：
大量データから
知識を抽出できる

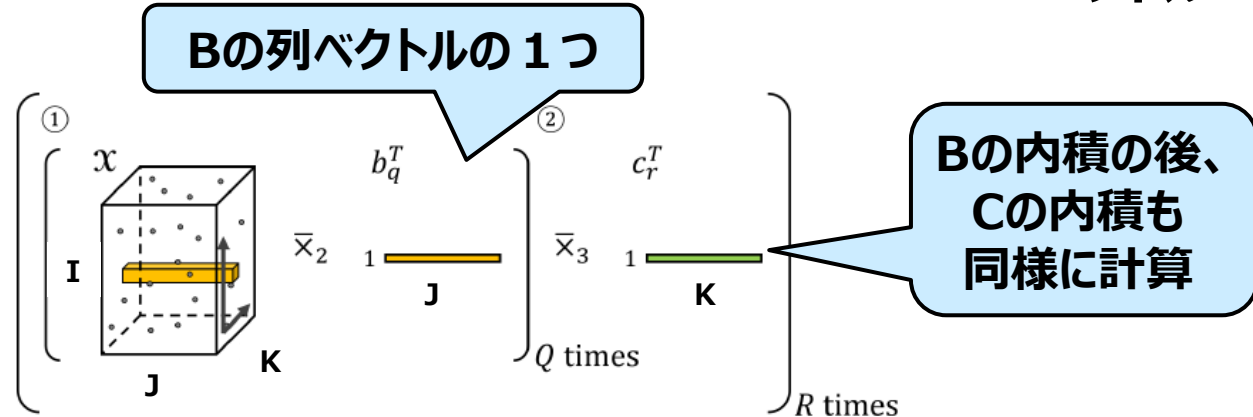


MapReduceによる並列化の課題

テンソル分解の基本
アルゴリズムは
交互最小二乗法



内積計算を
MapReduce
により並列化
(HATEN2-Naive)



[Map]

Key: $I \times K$ のインデックス、Value: 各ベクトルの値

※Bの列ベクトルは複製して全てのKeyで送る

大きさ IJK

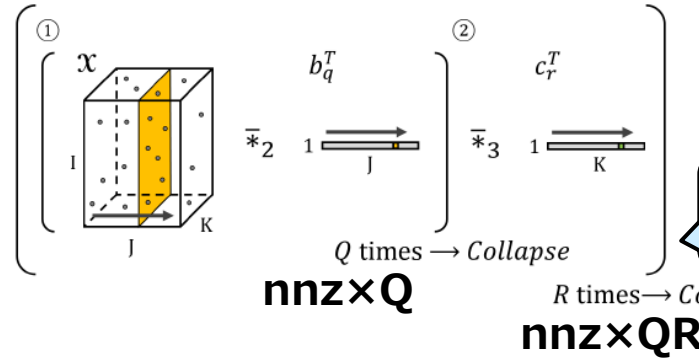
[Reduce]

ベクトルの内積を計算

課題：中間データが大きすぎる！（非零要素数+IJK）

提案：各ノードの処理を増やし中間データを減らす

(1) 行列の列ベクトルの
同じ要素に対応
する積は1つのノード
で計算
(HATEN2-DNN)

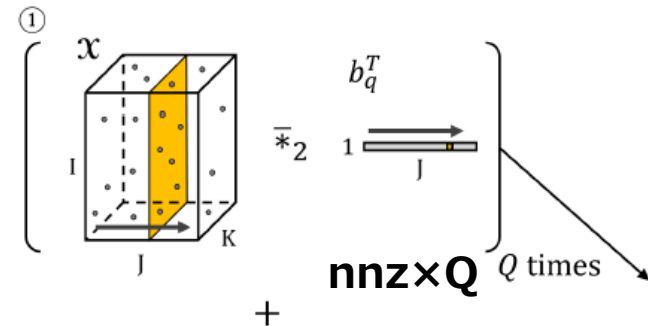


[MR1] 要素ごとの積
[MR2] 積を足して内積に
これらを行列数だけ行う

中間データはまだ大きい
(非零要素数) \times QR

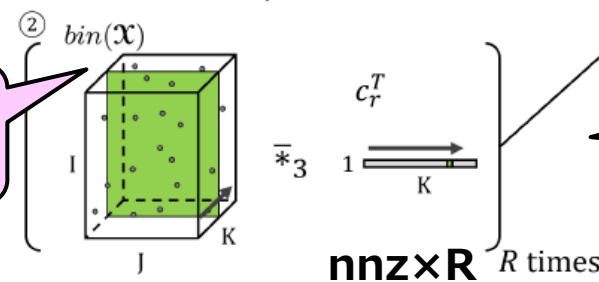
最終生成データ $I \times QR$
よりとても大きい!

(2) 各行列に対応する
積を計算しておいて
一気に足し合わせる
(HATEN2-DRN)



[MR1] 要素ごとの積
を行列数だけ行ってから
[MR2] 積を足し合わせる

注：1つ以外の行列
に対してはバイナリ化



中間データを大きく削減
(非零要素数) \times (Q+R)

メモリが大きいときに
処理数を減らす

(2') 行列の同じ行の要素の計算を1つのノードで行う(HATEN2-DRI)

評価実験

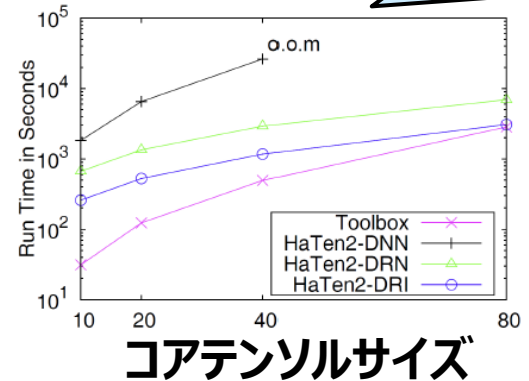
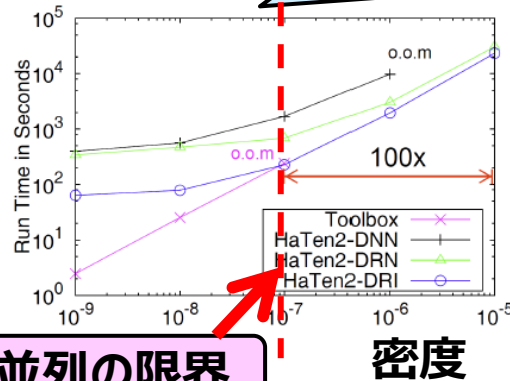
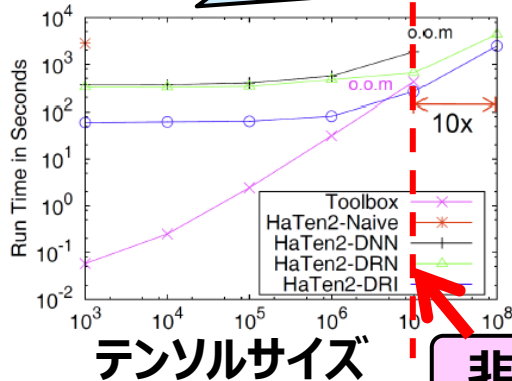
非並列では不可能な大量・高密度のテンソルを大コアサイズで分解可能

テンソルのサイズが大きくなっても大丈夫

テンソルの密度が大きくなっても大丈夫

コアサイズが大きくてもDRIなら高速

計算時間



非並列の限界

大規模データから知識抽出に成功(23M×23M×166、非零99M個)

コアテンソルの各要素に対応する概念

Concepts	Subject Entity	Object Entity	Relation
Concept1:(S2,O1,R2) "Instrumental Pieces"	Guitar	Faust: Soldatenchor	'ns:music.instrument.instrumentalists'
	Keyboard	3 Trios for Flute, Violin, Cello, no. 1: IV. Rondo: Allegro	'ns:music.instrument.variation'
	Drums	Piano Concerto in A minor, op. 54: III. Allegro vivace	'ns:music.album.artist'
Concept2:(S1,O1,R3) "Classics"	London Symphony Orchestra	Faust: Soldatenchor	'ns:music.release.region'
	Wolfgang Amadeus Mozart	3 Trios for Flute, Violin, Cello, no. 1: IV. Rondo: Allegro	ns:music.record-label.artist'
	Ludwig van Beethoven	Piano Concerto in A minor, op. 54: III. Allegro vivace	'ns:music.album.artist'
Concept3:(S3,O3,R3) "Musics from Labels"	Columbia	Sikidim (by Warner Music Group)	'ns:music.release.region'
	EMI	Terrifying Tales (by EMI)	'ns:record-label.artist'
	Atlantic Records	Rose of Tralee (by EMI)	'ns:music.album.artists'