Document Number: P1444R0
Date: 2019-01-21
Authors: Michael Wong
Project: Programming Language C++, SG19 Machine Learning
Reply to: Michael Wong <michael@codeplay.com>

SG19: Machine Learning 2018/12/14-2019/01/11

Contents

Minutes for 2018/12/14 SG19 Conference Call

1.1 Roll call of participants

David Lindelof, Romain Biessy, Sarthak Pati, John Lawson, Torvald Riegal, Ritwik Dubey, Vincent Reverdy, Michael Wong, Emad Barsoum, Matthieu Brucher, Peter Goldsborough,

1.2 Adopt agenda

Approve

1.3 Approve minutes from previous meeting, and approve publishing previously approved minutes to ISOCPP.org

Approve.

1.4 Action items from previous meetings

2. Main issues (125 min)

2.1 General logistics

2 new SGs mailing lists for SG19 Machine Learning

https://groups.google.com/a/isocpp.org/forum/#!forum/sg19

and SG20 Education

https://groups.google.com/a/isocpp.org/forum/?fromgroups=#!forum/sg20

https://isocpp.org/std/forums

Any meeting rooms required for Kona?

EWG-I will meet Mon-Wed.
LEWG-I is penciled in for Mon-Thu, considering the outcome in San Diego.

SG12/WG23 will meet Wed-Fri - the last day being SG12-focused.

SG20 (education) would like to meet on Thursday (all day)

Suggest SG14 Friday Morning , SG19 afternoon.

Who is coming?

Vincent, Peter and Sebastien, Michael,

2.2 Paper reviews

 Feedback to other papers :
Linear Algebra, SG14
https://groups.google.com/a/isocpp.org/forum/#!forum/sg14

Scope:
linear algebra in SG14 is fundamental
Peter: Graphing is needed as well, make interactive C++ useful., ecosystem changes
David: online travel agent , expedia,, performance algorithm,
Emad: MS deeplearning team, improve training framework and improve perforamnce, CNTK,
pytorch, tensorflow, scalbility multiple backends, how you deploy it, prior in computer vision
John: Codeplay, neural network team
Matthieu: PI on python, hPC computing,
Ritwik: Ecegy, financial world, real time data
Romain: Codeplay, SYCL backend of tensorflow,
Sarthak:cancer imaging C++ more acessible
torvald: Redhat, platofrm enginering organizer
Vincent: French delegation, UUIC, Paris Obs, astrophysics, high performance data structures, to
speed up trees and ML
Michael Wong: CP VP interested in ML for self-driving cars

grahiing library: SG13
graphing expression, trees DAG,
MW: Swift GEP, Tensorflow
https://llvm.org/devmtg/2018-10/slides/Hong-Lattner-
SwiftForTensorFlowGraphProgramExtraction.pdf
Emad: Julia is also going that way
PG: are there any main interfaces that is interested
VR: as an expression, tree

Quantization: float16 and bfloat16, int 8,
Generalized floating point format
Automatic differentiation is likely Sg19,

anyone working with Jupyter notebook interactive C++? Does not work in many cases

Interactive C++?

Do we need something like Boost graph library? Each framework write their own, some
graphing library usually have more restrictions with data exchange between nodes is likely a
sparse tensor,

is this so complex that everyone is doing their own? CNTK graph has cycles in it,

dynamic graphs and static graph for most other cases:

Papers for Kona?
Feedback for LA for ML

https://docs.google.com/document/d/1poXfr7mUPovJC9ZQ5SDVM_1Nb6oYAXlK_d0ljdUAtSQ/edit
https://docs.google.com/viewer?a=v&pid=forums&srcid=MTE5NTAwNjk0ODI0NDg0MTc0MjkBMTIxOTM2MjE3MDIyMDkwMjA2NDgBS0NEbHlzaGZZDZ0FKATAuMQFpc29jHAub3JnAXYy&authuser=0

missing inner product,
AI: Mattheiu to list initial requirements from ML, track what is added to ML and what we need to do on our own

the next LA meeting is Jan 2, 1-3 ET

Feedback for FP16
- hide quoted text -

Next call: Jan 11

2.2.2 any other proposal for reviews?

2.3 Other Papers and proposals

2.5 Future F2F meetings:

2.6 future C++ Standard meetings:

https://isocpp.org/std/meetings-and-participation/upcoming-meetings

- **2019-02-18 to 23: Kona, HI, USA;** Standard C++ Foundation, NVIDIA, Plum Hall, Jens Maurer
- **2019-07-15 to 20: Cologne, Germany;** Nicolai Josuttis
- **2019-11-04 to 09: Belfast, Northern Ireland;** Archer Yates

3. Any other business
Reflector
https://groups.google.com/a/isocpp.org/forum/#!newtopic/sg19

Code and proposal Staging area

4. Review

4.1 Review and approve resolutions and issues [e.g., changes to SG's working draft]

4.2 Review action items (5 min)

5. Closing process

5.1 Establish next agenda

Jan 11

5.2 Future meeting

Dec 14: this meeting
Jan 11:
Jan 21: Kona mailing deadline
Feb 8:
Feb 18: C++ Std meeting Kona

Minutes for 2019/01/11 SG19 Conference Call

1.1 Roll call of participants

Michael, David Gillies, David Lindelof, Frank Seide, Guiherme Hartmann, Sebastien Messmer, Sylvain Corlay, Johan Mabille, John Lawson, Uwe Dolinski, Vincent

1.2 Adopt agenda

Approve

1.3 Approve minutes from previous meeting, and approve publishing previously approved minutes to ISOCPP.org

Approve

1.4 Action items from previous meetings

2. Main issues (125 min)

2.1 General logistics

2 new SGs mailing lists for SG19 Machine Learning

https://groups.google.com/a/isocpp.org/forum/#!forum/sg19

and SG20 Education

https://groups.google.com/a/isocpp.org/forum/?fromgroups=#!forum/sg20

https://isocpp.org/std/forums

Any meeting rooms required for Kona?

EWG-I will meet Mon-Wed.
LEWG-I is penciled in for Mon-Thu, considering the outcome in San Diego.

SG12/WG23 will meet Wed-Fri - the last day being SG12-focused.

SG20 (education) would like to meet on Thursday (all day)

SG14 Friday Morning, SG19 Friday afternoon.

Who is coming?

Frequency of meetings, we are actually booked weekly.

Mailing deadline is Jan 21
www.open-std.org/jtc1/sc22/wg21/docs/papers/2019/

2.2 Paper reviews

Any papers proposed for review at Kona? Deadline Jan 21 10 ET.

2.2.1 SG14 Linear Algebra progress Feedbacks
presentation by Matheiu
Different layers of proposal
https://docs.google.com/document/d/1poXfr7mUPovJC9ZQ5SDVM_1Nb6oYAXlK_d0ljdUAtSQ/edit

with dynamic pytorch graph, it is harder to be added to C++ apps
can use torch script? but lib torch is more flexible but still need to rewrite the code
restrict it to a smaller set for torch script

also need 3 or 4d arrays, need specialization for 4d arrays from ONYX
baseline shoudl enable unlimited or very large numbers in reality
ONYX was on purpose to expose accelerators from CUDNN, specfically, but had exceptions for higher dimensions
simialr in xtensor, the primary is n-dim arrays
some of the API are optmizmed to the backend in ONYX
can this be adapted to major exchange interface libraries like ONYX and NNEF, we should ensure Linear algebra can do this

this is a potential paper for Kona, write as a google doc

Jan 3 minutes

https://1195006948244841729.googlegroups.com/attach/60d4bfb6a587d/02jan2019.txt?part=0.1&view=1&vt=ANaJVrEPWoxHPpZ54t6CiYjZtaaNNv6rM23380AXmdr0gyxB6X5rLNiYkVM65h1TPpvpjZoqeaUgjRnsRUTe3ZVwfPvQyBQl27VnYlaN-y4j8EEcf0jWCWY

2.2.2 any other proposal for reviews?

xtensor presentation
expression template for n-d library, use numpy api, broadcast, reshape,
3 type of containers
integrates with Blas, or MKL,
also have adaptors, for old-style C arrays

https://github.com/QuantStack/xtensor
from Johan Mabille to everyone:
https://xtensor.readthedocs.io/en/latest/numpy.html
from Sylvain Corlay & Johan Mabille to everyone:
closure semantics
from Sylvain Corlay & Johan Mabille to everyone:
https://xtensor.readthedocs.io/en/latest/closure-semantics.html
from Sylvain Corlay & Johan Mabille to everyone:
https://mybinder.org/v2/gh/QuantStack/xtensor/stable?filepath=notebooks/xtensor.ipynb

needs list expression access by generalizing cref and ref, only one type that is closure, so when taking a ref, it is kept as a ref
so when concat, will expand the lifetime,
plan for gpu in future, already suppord SIMD using xsimd like boost:simd can be competitive with eigen; pyThran is a python to C++ translator is trying to make xtensor a backend to numpy

memory space is different in GPU, so needs to be copied and moved a lot, how to deal with that?
tensor added that is already in GPU may be loaded by batch,
but can you know if something is modified in oneside and reduce the copying
xFrame

sparse tensor? its an expression system, so can make the tensor for any data structure, CSD and CSR is in the plan
memory management? there are 2 : containers that we do provide are backed by std:vector
internally using your own allocator, then there is the lifetime management

Graphs

Probability

A prioritized layer approach to ML
http://www.open-std.org/jtc1/sc22/wg21/docs/papers/2018/p1360r0.pdf

fundamental arrays, matrix, vectors, tensors, linear algebra
●
facilitate better support for interchange of in memory information/data between packages
●
basic graphing
●
optimization , quantization, parallelism, batching computations of vector, matrix, tensors

- packaging to allow adding computation/data manipulation/scaling packages + dependency
- supporting portability to various hardware embedded inference engines, and up down convert of different FP sizes between training and inference.
- support of exchange formats (ONNX, NNEF)
- support for kernel fusion on training and inference
- support of accelerator dispatch to inference engines, GPUs, FPGAs, MPSoC, Tensor Processing Units,

  Coarse Grain Reconfigurable Arrays

  , many of the newer ML boards from Xilinx, Google, ARM, Wave Computing, Nvidia
- do we need a graph extraction pass on top of C++
- support interoperability with data formats from Python and R packages
- Other C++ ML/Data Analysis libs : shark, MLpack, dlib, root [

  root

  ]
- Integration Xla, tvm, tensor-rt, glow
- lazy evaluation execution graphs/workflows
- graph and tree data structures

2.3 Other Papers and proposals

2.5 Future F2F meetings/Conferences:

- 2019-06-10-15 ICML Long Beach US
- 2019-06-15-21: CVPR Long Beach US

- 2019-10-27-11-3 ICCV Seoul

2.6 future C++ Standard meetings:

https://isocpp.org/std/meetings-and-participation/upcoming-meetings

- **2019-02-18 to 23: Kona, HI, USA;** Standard C++ Foundation, NVIDIA, Plum Hall, Jens Maurer
- **2019-07-15 to 20: Cologne, Germany;** Nicolai Josuttis
- **2019-11-04 to 09: Belfast, Northern Ireland;** Archer Yates

2.7 Other Standard activities
Khronos ML
https://www.khronos.org/machine-learning/
ISO SC 42 AI
https://www.iso.org/committee/6794475.html

3. Any other business
Reflector
https://groups.google.com/a/isocpp.org/forum/#!newtopic/sg19

Code and proposal Staging area

4. Review

4.1 Review and approve resolutions and issues [e.g., changes to SG's working draft]

4.2 Review action items (5 min)

5. Closing process

5.1 Establish next agenda
Graphs + prep Kona

Feb 8

5.2 Future meeting

Dec 14:logistics
Jan 11:this meeting
Jan 21: Kona mailing deadline
Feb 8:
Feb 18: C++ Std meeting Kona